

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 27, 2003, 18:20:16 ; Search time 25 seconds

(without alignments)  
944.296 Million cell updates/sec

Title: US-10-088-045-2

Perfect score: 2626  
Sequence: 1 MSKPLIKITLLICALSALML.....NRGGLSADDAKAKKTKPN 502

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1386	52.8	392	1 ANIA_NEIGO	002219 neisseria g
2	1468	17.8	374	1 NIR_RHOSH	053239 rhodobacter
3	461.5	17.6	378	1 NIR_ACHCY	P25006 achromobact.
4	453	17.3	379	1 NIR_RHIGA	001537 rhizobium g
5	451	17.2	363	1 NIR_PSECL	006006 pseudomonas
6	439.5	16.7	376	1 NIR_RHIME	092229 rhizobium m
7	437	16.6	377	1 NIR_RHIME	060214 rhizobium h
8	427.5	16.3	376	1 NIR_ALCPA	P38501 alcaligenes
9	424	16.1	330	1 NIR_ALCPX	P81445 alcaligenes
10	203.5	7.7	605	1 PCOA_ECOLI	047452 escherichia
11	170	6.5	552	1 ASO_CUCPM	P37064 cucurbita p
12	167	6.4	579	1 ASO_CUCMA	P24792 cucurbita m
13	164.5	6.3	609	1 CPAL_PSESM	P12374 pseudomonas
14	163	6.2	478	1 CYCA_GLUOX	047594 gluconobact
15	162	6.2	578	1 ASO_TOBAC	040588 nicotiana t
16	161	6.1	148	1 C552_THETH	P04164 thermus the
17	157.5	6.0	589	1 CPAA_PSESM	P59571 pseudomonas
18	153.5	5.8	587	1 ASO_CUCSA	P14133 cucumis sat
19	146	5.6	468	1 CYCA_ACEPO	003318 acetoabacter
20	145	5.5	555	1 ASO_BRANA	000624 brassica na
21	143.5	5.5	2211	1 FAS_BOVIN	Q28107 bos taurus
22	137	5.2	2258	1 FAS_PIG	099191 sus scrofa
23	133.5	5.1	576	1 LACI_THACU	P56193 thatephor
24	131	5.0	531	1 LAC4_THACU	002081 thatephor
25	128.5	4.9	2224	1 FAS_HUMAN	P12529 homo sapien
26	124	4.7	533	1 CUBO_YERPE	082600 yerisina pe
27	124	4.7	572	1 LAC2_THACU	P02079 thatephor
28	118	4.4	520	1 LAC1_YEAST	P43561 saccharomyc
29	116	4.4	520	1 LAC1_AGABI	012541 agaricus bi
30	116	4.4	652	1 NOS2_PARDE	051705 paracoccus
31	111.5	4.2	554	1 NTP2_TOBAC	P29162 nicotiana t
32	111	4.2	527	1 LAC3_TRAVI	099056 trameetes vi
33	111	4.2	1101	1 GUNC_CELFI	P14090 cellulomona

34	109	4.2	108	1 C555_CHLIT	08X93 chlorobium
35	109	4.2	1062	1 CERU_MOUSE	Q61147 mus musculu
36	108	4.1	527	1 LACS_TRAVE	Q12717 trameetes ve
37	107.5	4.1	520	1 LAC2_AGABI	012542 agaricus bi
38	107.5	4.1	1257	1 CCAA_BACTU	045754 bacillus th
39	107.5	4.1	2647	1 FLNA_HUMAN	P21333 homo sapien
40	107	4.1	1654	1 OMPB_RICRI	Q53047 r outer mem
41	106.5	4.1	485	1 IMDB_PYRAB	Q9UY49 pyrococcus
42	106.5	4.1	599	1 LAC2_THACU	002075 thatephor
43	106.5	4.1	642	1 PHSA_STRAT	051692 streptomyc
44	105.5	4.0	739	1 PURL_CAUCR	Q9A5F0 caulobacter
45	105	4.0	86	1 C555_CHLIT	P00123 chlorobium

## ALIGNMENTS

RESULT 1  
ID ANIA\_NEIGO STANDARD; PRT; 392 AA.  
AC 002219;  
DT 01-JUN-1994 (Rel. 29, Created)  
DT 01-JUN-1994 (Rel. 29, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Major outer membrane protein Pan 1 precursor.  
OS ANIA.  
OS Neisseria gonorrhoeae.  
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;  
OC Neisseriaceae; Neisseria.  
OX NCBI\_TaxID=485;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN=RI0;  
RX MEDLINE=93014187; PubMed=1383156;  
RA Hoehn G.T., Clark V.L.;  
RT "Isolation and nucleotide sequence of the gene (ania) encoding the  
RT major anaerobically induced outer membrane protein of Neisseria  
RT gonorrhoeae.";  
RL Infect. Immun. 60:4704-4708(1992).  
RN (2)  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 33084 / F62;  
RX MEDLINE=93014188; PubMed=1398981;  
RA Hoehn G.T., Clark V.L.;  
RT "The major anaerobically induced outer membrane protein of Neisseria  
RT gonorrhoeae, Pan 1, is a lipoprotein.";  
RL Infect. Immun. 60:4704-4708(1992).  
CC -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid  
CC anchor (Probable).  
CC -1- SIMILARITY: Contains 2 plactocyanin-like domains.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; M97926; AAA25462.1; -  
CC PIR; A49208; A49208.  
CC PDB; 1KXV; 27-FEB-02.  
CC PDB; 1KXW; 27-FEB-02.  
CC InterPro; IPR00117; Cu-oxidase.  
CC InterPro; IPR001287; CuNO2\_reductase.  
CC Pfam; PF00394; Cu-oxidase; 2.  
CC PRINTS; PR00695; CUNO2RDTASE.  
CC PROSITE; PS00013; PROKAR\_LIPOPROTEIN; 1.  
KW Repeat; Outer membrane; Lipoprotein; Signal; 3D-structure.  
FT SIGNAL 1 18  
FT CHAIN 19 392 MAJOR OUTER MEMBRANE PROTEIN PAN 1.  
FT LIPID 19 19 N-ACYL DIGLYCERIDE (PROBABLE).  
FT

FT DOMAIN 101 195 PLASTOCYANIN-LIKE 1.  
 FT DOMAIN 245 346 PLASTOCYANIN-LIKE 2.  
 FT DOMAIN 368 387 4 X 5 AA TANDEM REPEATS OF A-A-S-A-P.  
 FT REPEAT 368 372 1.  
 FT REPEAT 373 377 2.  
 FT REPEAT 378 382 3.  
 FT REPEAT 383 387 4.  
 SO SEQUENCE 392 AA; 40954 MW; A4707CCB923C97 CRC64;

Query Match 52.8%; Score 1386; DB 1; Length 392;  
 Best Local Similarity 69.5%; Pred. No. 4,4e-90;  
 Matches 264; Conservative 44; Mismatches 66; Indels 6; Gaps 3;

11 LICALALMLSGSNQADKAPKSSIVDAAKTA-NADVAASOEHOGEPLVDATVTA 69  
 9 MIALSLPALAACG---GEQAQAQPAETPAASASASAAQATMETAGELPVIDATVTA 64  
 70 PEVPPVDRDHPAKVAVVKKETVEKVRRLADGVEYQFWTEGGVPGQMIRREGDTIEVQF 129  
 65 PEVPPAIDRDYPAKVAVVKKETVEKTKMDGVEYRWTFPGDVPGMIRREGDTIEVER 124  
 130 SNHPDSKMPHNVDPHAATGPGGAASFTAPGHTSTFSFKALQPLGYVHCAPVGMH 189  
 125 SNHPSTVPHNVDPHAATGPGGAATFTAPGRTSTFSFKALQPLGYVHCAPVGMH 184  
 190 ANGMVGLIIVEPEKGPVKVDKEYVMQGDFTYTKGKGEOGLQEPDEKAIREDAEVVFN 249  
 185 ANGMVGLIIVEPEKGPVKVDKEYYIQGDDPTTGKKGAGLQEPDEKVAEPEVVFVN 244  
 250 GSVGALTGRENALKAVGETVRLFVNGGPNLTSSFHVIGIFDKVHEGKGNNHIQTT 309  
 245 GHVGSJAGDNVALAKAGETVRMYVNGGPNLVSSFHVIGIFDKVYVEGKLIENNVQST 304  
 310 LIPAGGAATTEFVNDVPGDVVLVDHAIFFAFNGKALGILKEEENHEIYSHQTDVAVL 369  
 305 IYPAGGSAIVEFVNDIPGSLTLVDHSLFFAFNGKALGILKEEENHEIYSHQTDVAVL 364  
 370 PEGAPQAITDQBAKTPAPAP 389  
 365 GSGAASH-PAASAPMAASAPA 383

RESULT 2  
 NIR\_RHOSH STANDARD; PRT; 374 AA.

AC 053239;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Copper-containing nitrite reductase precursor (EC 1.7.2.1) (Cu-NIR).  
 GN NIRK.  
 OS Rhodobacter alphaeroides (Rhodospseudomonas alphaeroides).  
 OC Bacteria; Rhodospirillum rubrum; Alphaproteobacteria; Rhodospirillales;  
 OC Rhodospirillales; Rhodospirillum.  
 OX NCBI\_TaxID=1063;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=2.4.3;  
 RX MEDLINE=971715533; PubMed=9023188;  
 RA Tosques I.E., Kwiatkowski A.V., Shi J., Shapleigh J.P.;  
 RT "Characterization and regulation of the gene encoding nitrite  
 reductase in Rhodobacter alphaeroides 2.4.3.";  
 RL J. Bacteriol. 179:1090-1095(1997).  
 CC -1- CATALYTIC ACTIVITY: Nitrite oxidase + H(2)O + ferricytochrome c =  
 nitrite + ferrocyanochrome c.  
 CC -1- COFACTOR: CONTAINS ONE ATOM OF TYPE 1 COPPER AND ONE ATOM OF TYPE  
 II COPPER; FAD. TYPE 1 COPPER IS BOUND WITHIN A SINGLE MONOMER,  
 WHILE TYPE 2 COPPER IS HELD BY RESIDUES FROM EACH OF TWO MONOMERS  
 OF THE TRIMER. NITRITE IS BOUND TO THE TYPE 2 COPPER SITE.  
 CC PSEUDAZURIN IS THE PHYSIOLOGICAL ELECTRON DONOR FOR THE CU-NIR IN  
 VITRO (BY SIMILARITY).  
 CC -1- PATHWAY: Nitrate assimilation (denitrification).  
 CC -1- SUBUNIT: Homotrimer (By similarity).

CC -1- SUBCELLULAR LOCATION: Periplasmic (By similarity).  
 CC -1- DOMAIN: THE TYPE 1 COPPER SITE IN NIR PLAYS A CRUCIAL ROLE FOR  
 CC ELECTRON TRANSFER FROM PSEUDAZURIN TO THE TYPE II COPPER SITE  
 CC OF NIR, WHICH COMPRISES THE CATALYTIC CENTER OF NIR FOR THE  
 CC REDUCTION OF NITRITE.  
 CC -1- SIMILARITY: Contains 2 plastocyanin-like domains.  
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

DR EMBL: U62291; AAB05767.1; -  
 DR HSSP: P25006; INIP.  
 DR InterPro: IPR001117; Cu-oxidase.  
 DR InterPro: IPR001287; CuNO2\_reductase.  
 DR Pfam: PF00394; Cu-oxidase; 2.  
 DR PRINTS: PR00695; CUNO2RDPAE.  
 KM Oxidoreductase; Copper; Metal-binding; Flavoprotein; FAD;  
 KW Nitrate assimilation; Repeat; Periplasmic; Signal.  
 FT CHAIN 1 38  
 FT SIGNAL 39 374  
 FT DOMAIN 93 189  
 FT PLASTOCYANIN-LIKE 1.  
 FT METAL 126 126  
 FT METAL 131 131  
 FT METAL 166 166  
 FT METAL 167 167  
 FT METAL 177 177  
 FT METAL 182 182  
 FT METAL 338 338  
 SQ SEQUENCE 374 AA; 40308 MW; 3406B58E7DD9934 CRC64;

Query Match 17.8%; Score 468; DB 1; Length 374;  
 Best Local Similarity 33.8%; Pred. No. 1.2e-25;  
 Matches 133; Conservative 52; Mismatches 128; Indels 80; Gaps 15;

QY 34 KSTVDAAA-----KTANDNNAASOEHOGEPLVDATVTAHAP-----VP 73  
 DB 5 RAALVGAAALASAPLIVIRTAGAQAAPAO-----LASAAPVDLSNLPVKHTLVP 53  
 QY 74 PVVDRDHPAKVAVVKKETVEKVRRLADGVEYQFWTEGGVPGQMIRREGDTIEVQF 126  
 DB 54 PEPANAEQVAAAGPVINEFEMRIIEKEVQLEDAVLQMTDGSIPGLMTVHESDYVE 113  
 QY 127 VQSNHPDSKMPHNVDPHAATGPGGAASFTAPGHTSTFSFKALQPLGYVHCAPVGMH 186  
 DB 114 LTLINPENTWPHNIDFHAATGALGGGLTLINPGEKVLRFKATRAAFVYHC-APG 171  
 QY 187 ---MHANGATGLIIVEPEKGP-----KVDKEYVMQGDFTYTKGKGEOGLQ 231  
 DB 172 PMIPMHVVSQMGAGICVLPDRGLKHEGKPVYVDVYVYIGESDHYIPKDEGTY-----M 226  
 QY 232 PFDMKAIRED-----AEVYVNGSVGALTGRENALKAVGETVRLFVNGGPNLTSS 282  
 DB 227 RSTSTPEGEYEDVAVVADTLIPSHIVFNGAVGALTGRENALKAVGDNV-LFV-HSQKRS 284  
 QY 283 SFHVIGEIFDKVHEGKGNNHIQTTLLPAGGAATTEFVNDVPGDVVLVDHAIFFA 339  
 DB 285 RHHLIGCHDVL-WEIGKFNHAPERNLETFIRGAGAAALVKFLQPGVAYAVNHLIEA 343  
 QY 340 ENKAGALGILKEEENHEIYSHQTDVAVLPEG 372  
 DB 344 VHKGATASVAVGEGWMDL-----MEQVMPVPG 371

RESULT 3  
 NIR\_ACHCY STANDARD; PRT; 378 AA.  
 ID NIR\_ACHCY  
 AC P25006;



DT 01-MAR-1992 (rel. 21, Created)

DT 01-NOV-1997 (rel. 35, Last sequence update)

DT 15-SEP-2003 (rel. 42, Last annotation update)

DE Copper-containing nitrite reductase precursor (EC 1.7.2.1) (Cu-NIR).

GN NMR.

OS Achromobacter cycloclastes.

OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;

OC Alcaligenaceae; Achromobacter.

OX NCBI\_TaxID=223;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=IAM 1013;

RX MEDLINE=96193667; PubMed=8605003;

RA Chen J.-Y., Chang W.-C., Chang T., Chang W.-C., Liu M.-Y., Payne W.J.,

RA le Gall J.;

RT "Cloning, characterization, and expression of the nitric oxide-

RT generating nitrite reductase and of the blue copper protein genes of

RT Achromobacter cycloclastes.";

RL Biochem. Biophys. Res. Commun. 219:423-428 (1996).

RN [2]

RP SEQUENCE OF 39-378.

RC STRAIN=IAM 1013;

RX MEDLINE=91308101; PubMed=1830217;

RA Penderson F.F., Kumar S., Adman E.T., Liu M.-Y., Payne W.J.,

RA le Gall J.;

RT "Amino acid sequence of nitrite reductase: a copper protein from

RT Achromobacter cycloclastes.";

RL Biochemistry 30:7180-7185 (1991).

RN [3]

RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).

RX MEDLINE=91320115; PubMed=1862344;

RA Godden J.W., Turley S., Teller D.C., Adman E.T., Liu M.-Y.,

RA Payne W.J., le Gall J.;

RT "The 2.3-A X-ray structure of nitrite reductase from Achromobacter

RT cycloclastes.";

RL Science 253:438-442 (1991).

RN [4]

RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).

RX MEDLINE=96070866; PubMed=749203;

RA Adman E.T., Godden J.W., Turley S.;

RT "The structure of copper-nitrite reductase from Achromobacter

RT cycloclastes at five pH values, with NO2-bound and with type II

RT copper depleted.";

RL J. Biol. Chem. 270:27458-27474 (1995).

CC -I- CATALYTIC ACTIVITY: Nitric oxide + H(2)O + ferricytochrome c =

CC nitrite + ferrocyclochrome c.

CC -I- COFACTOR: CONTAINS ONE ATOM OF TYPE I COPPER AND ONE ATOM OF TYPE

CC II COPPER, FAD.

CC -I- PATHWAY: Nitrate assimilation (denitrification).

CC -I- SUBUNIT: Homotrimer.

CC -I- SUBCELLULAR LOCATION: Periplasmic.

CC -I- DOMAIN: THE TYPE I COPPER SITE IN NIR PLAYS A CRUCIAL ROLE FOR

CC ELECTRON TRANSFER FROM PSEUDOURIN TO THE TYPE II COPPER SITE

CC OF NIR, WHICH COMPRISES THE CATALYTIC CENTER OF NIR FOR THE

CC REDUCTION OF NITRITE.

CC -I- SIMILARITY: Contains 2 plastocyanin-like domains.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation-

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL, Z48635; CAA88564.1; -

CC PIR, JC4648; JC4648.

DR PDB, 2NRD; 07-DEC-95.

DR PDB, 1N1A; 07-DEC-95.

DR PDB, 1N1B; 07-DEC-95.

DR PDB, 1N1C; 07-DEC-95.

DR PDB, 1N1D; 07-DEC-95.

DR PDB, 1N1E; 07-DEC-95.

DR PDB, 1N1F; 07-DEC-95.

DR InterPro; IPR001117; Cu-oxidase.

DR InterPro; IPR001287; CuNO2\_reductase.

DR InterPro; IPR006311; Tat.

DR Pfam; PF00394; Cu-oxidase; 2.

DR PRINTS; PR00695; CUNO2RPTASE.

DR TIGRFAMs; TIGR01409; Tat\_signal\_seq; 1.

KW Oxidoreductase; Copper; Metal-binding; Flavoprotein; FAD;

KW Nitrate assimilation; Repeat; Periplasmic; Signal; 3D-structure.

FT SIGNAL 1 38

FT CHAIN 39 378

FT DOMAIN 39 213

FT DOMAIN 214 378

FT METAL 133 133

FT METAL 138 138

FT METAL 173 173

FT METAL 174 174

FT METAL 183 183

FT METAL 188 188

FT METAL 344 344

FT HELIX 47 49

FT STRAND 52 54

FT TURN 60 61

FT STRAND 69 69

FT STRAND 76 76

FT TURN 92 93

FT STRAND 96 102

FT TURN 103 104

FT STRAND 105 105

FT TURN 110 114

FT STRAND 115 116

FT TURN 127 128

FT STRAND 133 133

FT STRAND 136 137

FT TURN 138 139

FT HELIX 143 149

FT STRAND 152 152

FT TURN 154 155

FT STRAND 156 163

FT STRAND 168 173

FT TURN 177 178

FT HELIX 180 184

FT TURN 185 187

FT STRAND 189 195

FT TURN 197 197

FT STRAND 198 198

FT TURN 200 201

FT STRAND 203 204

FT STRAND 207 208

FT STRAND 212 221

FT STRAND 225 225

FT TURN 227 228

FT STRAND 231 231

FT TURN 237 248

FT HELIX 249 250

FT STRAND 255 258

FT TURN 259 260

FT STRAND 261 261

FT TURN 262 265

FT HELIX 267 269

FT STRAND 271 274

FT TURN 275 276

FT STRAND 278 285

FT STRAND 290 290

FT STRAND 292 295

FT TURN 296 296

FT STRAND 299 303

FT TURN 304 305

FT STRAND 308 309

FT TURN 313 316

FT STRAND 321 321

FT TURN 323 324

FT STRAND 325 332  
 FT STRAND 337 343  
 FT HELIX 346 350  
 FT TURN 351 352  
 FT STRAND 355 361  
 FT TURN 366 368  
 SQ SEQUENCE 376 AA; 40771 MW; A70B52B814090EAS CRC64;

Query Match 17.6%; Score 461.5; DB 1; Length 378;  
 Best Local Similarity 36.3%; Pred. No. 3.3e-25;  
 Matches 134; Conservative 45; Mismatches 149; Indels 41; Gaps 12;

19 MUGSGNQDAKQAPKSTVDAAKTANNDNAASQHQGLPYPIDYIVTHAEVPPVVR 78  
 12 MLAGALAG--AVAPLHTAQAHAAGAAAAGAAPDITLPRVKDL-----VKPPFH 64  
 79 DH-----PAKVVMKETEVEKVMRL-ADGVYQFWTFGGGVPGQMIRVREGDTTEVOPS 130  
 65 AHDOYAKTGRPRVVEFTMTIEKKLVIDRGSTEIHAMTFNGSVPGPLMVHENDYVELRLI 124  
 131 NHPDSKMPHNVDFHAATGPGGAASFTAPGHTSTSFALOPGLVYHCAVAVGM--- 187  
 125 NEDTTLNIDIFHAATGALGGALTVQVNPGETTLRFATKPGVFVHC--AFEGMVPW 182  
 188 HANGMYGLIVEPKKGLP-----KVDKEYVVMQDFFYTK-----GKYGEQGLQEP 233  
 183 HTSGNGAIWVLPFRDGLDDEKQPLTYDKIYVGSQDFVPPDEKNGYKKTETPEAYE 242  
 234 DMEKAIRE-DAEYVFNQSVGALTGENALKAKEGTVRLFEVNGGPNLTSSFFHVEIFD 292  
 243 DAVKAMRTLTPTPHIVNGAVGALTGDHALTAAGE--RVLVHSAQNRDTRPHLIGHD 300  
 293 KVHEFGKGEKNH--NIOTTLIPAGGAATTEKVDVPGDYVLVDHAFPAFKGALGILK 349  
 301 YV-WATGKFRNPEDLDQETWLTIPGTAGAAFYTFRQPGVAYVNNLIEAFELGAGHKK 359  
 350 VEGEENHEI 358  
 360 VTGEWMDL 368

RESULT 4  
 NLR\_RHIGA STANDARD; PRT; 379 AA.

AC 001537;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Copper-containing nitrite reductase precursor (EC 1.7.2.1) (Cu-NIR).  
 GN NITR.  
 OS Rhizobium galegae.  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 OC Rhizobiaceae; Rhizobium/Agrobacterium group; Rhizobium.  
 OX NCBI\_TaxId=399;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C-179;  
 RX MEDLINE=93175864; PubMed=8439151;  
 RA Ye R.W.; Fries M.R.; Bezborodnikov S.G.; Averill B.A.; Tiedje J.M.;  
 RT "Characterization of the structural gene encoding a copper-containing  
 RT nitrite reductase and homology of this gene to DNA of other  
 RT denitrifiers." Microbiol. 59:250-254(1993).  
 RL Appl. Environ. Microbiol. 59:250-254(1993).  
 CC -1- CATALYTIC ACTIVITY: Nitric oxide + H(2)O + ferricytochrome c =  
 CC nitrite + ferrocyclochrome c.  
 CC -1- COFACTOR: CONTAINS ONE ATOM OF TYPE I COPPER AND ONE ATOM OF TYPE  
 CC II COPPER. FAD. TYPE I COPPER IS BOUND WITHIN A SINGLE MONOMER,  
 CC WHILE TYPE 2 COPPER IS HELD BY RESIDUES FROM EACH OF TWO MONOMERS  
 CC OF THE TRIMER. NITRITE IS BOUND TO THE TYPE 2 COPPER SITE.  
 CC PSEUDOURIN IS THE PHYSIOLOGICAL ELECTRON DONOR FOR THE CU-NIR IN  
 CC VITRO (BY SIMILARITY).  
 CC -1- PATHWAY: Nitrate assimilation (denitrification).  
 CC -1- SUBUNIT: Homotrimer (By similarity).

CC -1- SUBCELLULAR LOCATION: Periplasmic.  
 CC -1- DOMAIN: THE TYPE I COPPER SITE IN NIR PLAYS A CRUCIAL ROLE FOR  
 CC ELECTRON TRANSFER FROM PSEUDOURIN TO THE TYPE II COPPER SITE  
 CC OF NIR, WHICH COMPRISES THE CATALYTIC CENTER OF NIR FOR THE  
 CC REDUCTION OF NITRITE.  
 CC -1- SIMILARITY: Contains 2 plastocyanin-like domains.  
 CC -1- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-12 IS THE INITIATOR.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

DR EMBL; M97294; AAC79132.1; -;  
 DR PIR; A48936; A48936.  
 DR HSSP; P25006; NIF.  
 DR InterPro; IPR001117; Cu-oxidase.  
 DR InterPro; IPR001287; CUNO2\_reductase.  
 DR InterPro; IPR006311; Tat.  
 DR Pfam; PF00394; Cu-oxidase; 2.  
 DR PRINTS; PR00695; CUNO2RDTASE.  
 DR TIGRFAMs; TIGR01409; Tat signal seq; 1.  
 DR Trifam; TIGR01409; Tat signal seq; 1.  
 KW Oxidoreductase; Copper; Metal-binding; Flavoprotein; FAD;  
 KW Nitrate assimilation; Repeat; Periplasmic; Signal.  
 FT SIGNAL 1 32  
 FT CHAIN 33 379  
 FT DOMAIN 33 214  
 FT DOMAIN 215 379  
 FT METAL 134 134  
 FT METAL 139 139  
 FT METAL 174 174  
 FT METAL 175 175  
 FT METAL 184 184  
 FT METAL 189 189  
 FT METAL 345 345  
 SQ SEQUENCE 379 AA; 40694 MW; 090A3CBF66262F CRC64;

Query Match 17.3%; Score 453; DB 1; Length 379;  
 Best Local Similarity 37.3%; Pred. No. 1.3e-24;  
 Matches 118; Conservative 45; Mismatches 119; Indels 34; Gaps 10;

QY 72 VPPVDRDH-----PAKVVMKETEVEKVMRLAD-GVEYQFWTFGGGVPGQMIRVREGD 123  
 DB 59 VKPPVHAHTQKAESEPKVVERKMTTQEKIVDDKGTVEHAMTFDGSVPGPMIYHQDD 118  
 QY 124 TIEVQFSNHPDSKMPHNVDFHAATGPGGAASFTAPGHTSTSFALOPGLVYHCAVA 183  
 DB 119 YVELTLVNPDTNELQHNIDFHSATGALGGALTVNPGDTAVLRFRATKAGVFVHC--A 176  
 QY 184 PVGM--HANGMYGLIVEPKKGLP-----VDKEYVVMQDFFYTK-----GKYG 226  
 DB 177 PAGMVPWHTSGMNGAIWVLPFRDGLDKHGHLELVYDKVYVEODFYVPEKNGKFKYE 236  
 QY 227 EQLQEPDEKAIRE-DAEYVFNQSVGALTGENALKAKEGTVRLFEVNGGPNLTSSFFH 285  
 DB 237 SAGEAVPDVLEAMKMTLTPTPHIVNGAVGALTGDHALTAAGE--RVLIIHSAQNRDTRH 294  
 QY 286 VIGEIFDKVHFGKGEKNH--NIOTTLIPAGGAATTEKVDVPGDYVLVDHAFPAFK 342  
 DB 295 LIGCHODV-WATGKFRNPEDLDQETWLTIPGAGAAAYTFPOPGIYAVVNNLIEAFEL 353  
 QY 343 GALGILKVEEENHEI 358  
 DB 354 GAAGHKVYGDWMDL 369

RESULT 5  
 NLR\_PSECL STANDARD; PRT; 363 AA.  
 ID NLR\_PSECL  
 AC Q06006;

DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Copper-containing nitrite reductase precursor (EC 1.7.2.1) (Cu-NIR).  
 GN NIKR.  
 OS Pseudomonas chlororaphis (Pseudomonas aureofaciens).  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
 OC Pseudomonadaceae; Pseudomonas.  
 ON NCBI\_TaxID=333;  
 RX MEDLINE=93356602; PubMed=8352648;  
 RC STRAIN=ATCC 13985;  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 25-41.  
 RT "Copper-containing nitrite reductase from Pseudomonas aureofaciens is functional in a mutationally cytochrome cdi-free background (Nirs-)"  
 RT of Pseudomonas stutzeri.";  
 RU Arch. Microbiol. 160:18-26(1993).  
 CC -1- CATALYTIC ACTIVITY: Nitric oxide + H(2)O + ferricytochrome c =  
 CC nitrite + ferrocyclochrome c.  
 CC -1- COPFACTOR: CONTAINS ONE ATOM OF TYPE I COPPER AND ONE ATOM OF TYPE  
 CC II COPPER; FAD. TYPE I COPPER IS BOUND WITHIN A SINGLE MONOMER,  
 CC WHILE TYPE 2 COPPER IS HELD BY RESIDUES FROM EACH OF TWO MONOMERS  
 CC OF THE TRIMER. NITRITE IS BOUND TO THE TYPE 2 COPPER SITE.  
 CC PSEUDODAZURIN IS THE PHYSIOLOGICAL ELECTRON DONOR FOR THE CU-NIR IN  
 CC VITRO.  
 CC -1- PATHWAY: Nitrate assimilation (denitrification).  
 CC -1- SUBUNIT: Homotrimer (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Periplasmic.  
 CC -1- DOMAIN: THE TYPE I COPPER SITE IN NIR PLAYS A CRUCIAL ROLE FOR  
 CC ELECTRON TRANSFER FROM PSEUDODAZURIN TO THE TYPE II COPPER SITE  
 CC OF NIR, WHICH COMPRISES THE CATALYTIC CENTER OF NIR FOR THE  
 CC REDUCTION OF NITRITE.  
 CC -1- SIMILARITY: Contains 2 plastocyanin-like domains.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; 221945; CAA7939.1; -  
 CC DR PIR; S32112; S32112.  
 CC DR HSSP; P38501; 1A58.  
 CC DR InterPro; IPR001117; Cu-oxidase.  
 CC DR InterPro; IPR001287; CUNO2\_reductase.  
 CC DR Pfam; PF00394; Cu-oxidase; 2.  
 CC DR PRINTS; PR00695; CUNO2RDTASE.  
 CC KM Oxidoreductase; Copper; Metal-binding; Flavoprotein; FAD;  
 CC Nitrate assimilation; Repeat; Periplasmic; Signal.  
 CC KW Nitrate assimilation; Repeat; Periplasmic; Signal.  
 CC FT CHAIN 1 24  
 CC FT DOMAIN 25 363  
 CC FT DOMAIN 25 193  
 CC FT METAL 194 363 PLASTOCYANIN-LIKE 1.  
 CC FT METAL 113 113 COPPER (TYPE 1) (BY SIMILARITY).  
 CC FT METAL 118 118 COPPER (TYPE 2) (BY SIMILARITY).  
 CC FT METAL 153 153 COPPER (TYPE 2) (BY SIMILARITY).  
 CC FT METAL 154 154 COPPER (TYPE 1) (BY SIMILARITY).  
 CC FT METAL 163 163 COPPER (TYPE 1) (BY SIMILARITY).  
 CC FT METAL 168 168 COPPER (TYPE 1) (BY SIMILARITY).  
 CC FT METAL 324 324 COPPER (TYPE 2) (BY SIMILARITY).  
 CC SEQUENCE 363 AA; 39248 MW; 50DDB60CCAD3CB00 CRC64;  
 QY Query Match 17.2%; Score 451; DB 1; Length 363;  
 QY Best Local Similarity 39.1%; Pred. No. 1.7e-24;  
 QY Matches 124; Conservative 41; Mismatches 116; Indels 36; Gaps 14;  
 QY 72 VPPPPDRHH-----PAKVV-VKMETVEKVMPLAD-GYEYQWRTGGGVPGGMIRVRGCD 123  
 QY 38 VAPPLVHPHBOVSGPAPKVVQFRMSIEEKRWVIDDQGTLLQMTENGSGPFTLVVHSGD 97

QY 124 TIEVQFSNHPDPSKMHNDVFNHAATPGGCAEASFTAPGHTSTFSFKALQPLGYVHCAYA 183  
 DB 98 YIELTVLPATNSMHNVDVFNHAATGAGATGVVPCQGEVLRKADRSSTGFVYHC--A 155  
 QY 164 PYGM--HIANGMYGLILVERKEG--PK-----VDKEYVMQGFY-TKGKYGQGLQ 232  
 DB 156 PGQWPMVHVSGMNGALVLPDGDGRPOGKLLHYDRVYTTIGESDLTYPKDGHYKQVP 215  
 QY 233 FDMKAIKE-----DAEYVFNFGSVGALTGENALAKVGETVRLFPVNGCPNLTSGF 284  
 DB 216 -DLASVQDTRVAMTTLTPSHVFNFGVGTALGNALTSKGESEV-LPI-HQANDRSIP 272  
 QY 265 HVIGEIPDKVHEEGKGEN---HNIQTLLPAGGAITEFKVDVGDVYLVDAHAFRAN 341  
 DB 273 HLGHGGMV-WTTSKFPANPPQRMNETWFIQGSVAVALYFKQGTVYLSHNLIEANE 331  
 QY 342 KGALGILVVEGHEHET 358  
 DB 332 LGALAQIVREGQWDDDL 348  
 RESULT 6  
 NIKR\_RHIME STANDARD; PRT; 376 AA.  
 ID NIKR\_RHIME  
 AC Q92229;  
 DT 15-SEP-2003 (Rel. 42, Created)  
 DT 15-SEP-2003 (Rel. 42, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Copper-containing nitrite reductase precursor (EC 1.7.2.1) (Cu-NIR).  
 GN NIKR OR RA0681 OR SWA1250.  
 OS Rhizobium meliloti (Sinorhizobium meliloti).  
 OC Plasmid pSyma (megaplasmid 1).  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.  
 ON NCBI\_TaxID=382;  
 RX MEDLINE=21396509; PubMed=11481432;  
 RC STRAIN=1021;  
 RA Barnett M.J., Fisher R.F., Jones T., Komp C., Abola A.P.,  
 RA Barloy-Hubler F., Bowser L., Capela D., Galibert F., Gouzy J.,  
 RA Gurjal M., Hong A., Huizar L., Hyman R.W., Kahn D., Kahn M.L.,  
 RA Kalman S., Keating D.H., Palm C., Peck M.C., Surzycki R., Wells D.H.,  
 RA Yeh K.-C., Davis R.W., Federspiel N.A., Long S.R.;  
 RT "Nucleotide sequence and predicted functions of the entire  
 RT Sinorhizobium meliloti pSyma megaplasmid";  
 RT Proc. Natl. Acad. Sci. U.S.A. 98:9883-9888(2001).  
 CC -1- CATALYTIC ACTIVITY: Nitric oxide + H(2)O + ferricytochrome c =  
 CC nitrite + ferrocyclochrome c.  
 CC -1- COPFACTOR: CONTAINS ONE ATOM OF TYPE I COPPER AND ONE ATOM OF TYPE  
 CC II COPPER; FAD. TYPE I COPPER IS BOUND WITHIN A SINGLE MONOMER,  
 CC WHILE TYPE 2 COPPER IS HELD BY RESIDUES FROM EACH OF TWO MONOMERS  
 CC OF THE TRIMER. NITRITE IS BOUND TO THE TYPE 2 COPPER SITE.  
 CC PSEUDODAZURIN IS THE PHYSIOLOGICAL ELECTRON DONOR FOR THE CU-NIR IN  
 CC VITRO (BY SIMILARITY).  
 CC -1- PATHWAY: Nitrate assimilation (denitrification).  
 CC -1- SUBUNIT: Homotrimer (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Periplasmic (By similarity).  
 CC -1- DOMAIN: THE TYPE I COPPER SITE IN NIR PLAYS A CRUCIAL ROLE FOR  
 CC ELECTRON TRANSFER FROM PSEUDODAZURIN TO THE TYPE II COPPER SITE  
 CC OF NIR, WHICH COMPRISES THE CATALYTIC CENTER OF NIR FOR THE  
 CC REDUCTION OF NITRITE.  
 CC -1- SIMILARITY: Contains 2 plastocyanin-like domains.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AE007256; AAK65339.1; -

DR PIR. A95347; A95347.  
 DR InterPro; IPR001117; Cu-oxidase.  
 DR InterPro; IPR001287; CuNO2\_reductase.  
 DR InterPro; IPR006311; Tat.  
 DR Pfam; PF00394; Cu-oxidase; 2.  
 DR PRINTS; PR00695; CUNO2RDPASE.  
 DR TIGRFAMs; TIGR01409; Tat\_signal\_seg; 1.  
 DR OXIDOREDUCTASE; Copper; Metal-binding; Flavoprotein; FAD;  
 KM Nitrate assimilation; Repeat; Periplasmic; Signal; Plasmid;  
 KM Complete proteome.  
 FT SIGNAL 1 33 POTENTIAL.  
 FT CHAIN 34 376 COPPER-CONTAINING NITRITE REDUCTASE.  
 FT DOMAIN 98 193 PLASTOCYANIN-LIKE 1.  
 FT DOMAIN 258 359 PLASTOCYANIN-LIKE 2.  
 FT METAL 131 131 COPPER (TYPE 1) (BY SIMILARITY).  
 FT METAL 136 136 COPPER (TYPE 2) (BY SIMILARITY).  
 FT METAL 171 171 COPPER (TYPE 1) (BY SIMILARITY).  
 FT METAL 172 172 COPPER (TYPE 1) (BY SIMILARITY).  
 FT METAL 181 181 COPPER (TYPE 1) (BY SIMILARITY).  
 FT METAL 186 186 COPPER (TYPE 1) (BY SIMILARITY).  
 FT METAL 342 342 COPPER (TYPE 2) (BY SIMILARITY).  
 SQ SEQUENCE 376 AA; 40258 MW; 98799273001DC63 CRC64;  
 Query Match 16.7%; Score 439.5; DB 1; Length 376;  
 Best Local Similarity 33.2%; Pred. No. 1.7e-23;  
 Matches 121; Conservative 52; Mismatches 128; Indels 63; Gaps 13;  
 60 PVIDALVTTHAPF-----VPPVDRDHP-----AKVYVMEVKEGLPK- 92  
 24 PLIGAVSAABEAVAKTAHINVASLPRVKVDLVKPFVHAHTOKAEGSPKVEFTLTIES 83  
 93 -KVMRLADGEVYQFTFGGQVPCQMIRVREGDTIEVQFNSHPDSKPHNVDFAATPGG 151  
 84 KKIIVDEOGELHMTFNGSVPGPLVWVHDDVEVLTLPDNTNLOHNDIFHSATGALG 143  
 152 GAASFTAPGHTSTFSFKALQPLGYVYHCAVAVGM---HIANGYGLIVREKEGLPK- 206  
 144 GGLAVTVNPGDITVLAFFKSKAGAVFYHNC--APPGMVPWHTVSGMGALVWLPREGITDG 201  
 207 ----KVDEKYVMQGDYF---TKGRY-----GEOGLQPFMEKAIREDAEVYVNGS 251  
 202 KGNSTIYDKKYVYGEODFYVPRDANGKFKKYESGEAVADTLLEVMRTL--TPSHIVNGA 259  
 252 VGALTGBNALKAVGETVRLPVGNGCNLTSSRHVIGETIDKXHFEGKKEHN---NIQT 308  
 260 VGALTGSALKAAVGE--KVLIVHSQANRDTREHLTIGHGDYV-WATGKRKMAPVDQET 316  
 309 TLIPAGGAATTEPKVDVPGDYVLVDHAFRAFNKAGALGKVGGEENHEIYSHKQTDVAY 368  
 317 WFIPIGGTAGAAFYTFEOPGIYAYVNHNLIEAFELGAALAHFAVYGDNDLMT-----SVR 371  
 369 LPEG 372  
 372 APGG 375  
 Db 372 APGG 375  
 RESULT 7  
 NIR\_RHIHE STANDARD; PRT; 377 AA.  
 AC 060214;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Copper-containing nitrite reductase precursor (EC 1.7.2.1) (Cu-NIR).  
 GN NRK.  
 OS Rhizobium hedyarri.  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 OC Rhizobiaceae; Rhizobium/Agrobacterium group; Rhizobium.  
 OC NCBI\_TaxID=50338;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=HCNT1;  
 RX MEDLINE=97055705; PubMed=8899992;

RA Toffanin A., Wu Q., Maekus M., Casella S., Abruna H.D.,  
 RA Shapleigh J.P.;  
 RT "Characterization of the gene encoding nitrite reductase and the  
 RT physiological consequences of its expression in the nondifferentiating  
 RT Rhizobium 'hedyarri' strain HCNT1.";  
 RL Appl. Environ. Microbiol. 62:4019-4025(1996).  
 CC -1- CATALYTIC ACTIVITY: Nitric oxide + H(2)O + ferricytochrome c =  
 CC nitrite + ferrocyclochrome c.  
 CC -1- COFACTOR: CONTAINS ONE ATOM OF TYPE I COPPER AND ONE ATOM OF TYPE  
 CC II COPPER. FAD. TYPE I COPPER IS BOUND WITHIN A SINGLE MONOMER.  
 CC WHILE TYPE 2 COPPER IS HELD BY RESIDUES FROM EACH OF TWO MONOMERS  
 CC OF THE TRIMER. NITRITE IS BOUND TO THE TYPE 2 COPPER SITE.  
 CC PSEUDONITRIN IS THE PHYSIOLOGICAL ELECTRON DONOR FOR THE CU-NIR IN  
 CC VITRO (BY SIMILARITY).  
 CC -1- PATHWAY: Nitrate assimilation (denitrification).  
 CC -1- SUBUNIT: Homodimer (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Periplasmic (By similarity).  
 CC -1- DOMAIN: THE TYPE I COPPER SITE IN NIR PLAYS A CRUCIAL ROLE FOR  
 CC ELECTRON TRANSFER FROM PSEUDONITRIN TO THE TYPE II COPPER SITE  
 CC OF NIR, WHICH COMPRISES THE CATALYTIC CENTER OF NIR FOR THE  
 CC REDUCTION OF NITRITE.  
 CC -1- SIMILARITY: Contains 2 plastocyanin-like domains.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 CC EMBL; U65658; AAB05880.1; -.  
 DR HSSP; P38501; 1A57  
 DR InterPro; IPR001117; Cu-oxidase.  
 DR InterPro; IPR001287; CuNO2\_reductase.  
 DR InterPro; IPR006311; Tat.  
 DR Pfam; PF00394; Cu-oxidase; 2.  
 DR PRINTS; PR00695; CUNO2RDPASE.  
 DR TIGRFAMs; TIGR01409; Tat\_signal\_seg; 1.  
 DR OXIDOREDUCTASE; Copper; Metal-binding; Flavoprotein; FAD;  
 KM Nitrate assimilation; Repeat; Periplasmic; Signal.  
 KM Complete proteome.  
 FT SIGNAL 1 46 POTENTIAL.  
 FT CHAIN 47 377 COPPER-CONTAINING NITRITE REDUCTASE.  
 FT DOMAIN 99 194 PLASTOCYANIN-LIKE 1.  
 FT DOMAIN 259 360 PLASTOCYANIN-LIKE 2.  
 FT METAL 132 132 COPPER (TYPE 1) (BY SIMILARITY).  
 FT METAL 137 137 COPPER (TYPE 2) (BY SIMILARITY).  
 FT METAL 172 172 COPPER (TYPE 1) (BY SIMILARITY).  
 FT METAL 173 173 COPPER (TYPE 1) (BY SIMILARITY).  
 FT METAL 182 182 COPPER (TYPE 1) (BY SIMILARITY).  
 FT METAL 187 187 COPPER (TYPE 1) (BY SIMILARITY).  
 FT METAL 343 343 COPPER (TYPE 2) (BY SIMILARITY).  
 SQ SEQUENCE 377 AA; 40719 MW; C7AFEB3AE639E2F8 CRC64;  
 Query Match 16.6%; Score 437; DB 1; Length 377;  
 Best Local Similarity 31.3%; Pred. No. 1.7e-23;  
 Matches 123; Conservative 54; Mismatches 149; Indels 80; Gaps 14;  
 1 MSKPTLIKTLLIC-ALSALMLSGCSNQADKAQPKSITYDAAKTANADNAASQEHQGL 59  
 7 MTRRTVLTGAAVAGALTLPLITSSGGNASPTPVKLSATEIALAPRRKLD----- 55  
 60 PVIDALVTTHAPFPPVVDHP-----AKVYVMEVKEGLPK- 92  
 56 -----LVKPFVHVHTQKAEGRPVAVETLTIEKKLVIDGKTEVNAAMTFDGS 104  
 112 VPGMIRVREGDTIEVQFNSHPDSKPHNVDFAATPGGASFTAPGHTSTFSFKAL 171  
 105 VPGPLIVHDDVEVLTLPVNPENITLOHNDIFHSATGALGALTYVNPESAVLRKAT 164  
 172 QPGLVYVHCAVAVGM---HIANGYGLIVREKEGLPK-----VDKEYVMQGDYF- 220  
 165 KAGVIFYHC--APPGMVPWHTVSGMGALVWLPREGITDGKELVYDKVYVYGEODFYI 222

```

QY 221 ---TKG---KXGEGIQPFMEKAIRE-DAEYVFNQSGALGTENALAKVETVRLPV 273
DB 223 PRDEKGEFKKXDSFGCAVEYDVAVMRTLPTHTIVFNQAVGALGTENALFAAVE--RVL1 280
QY 274 GNGCPRLTSSFFHYIGISIFDKVHEGKGK-----NNIQTILIPAGAAITEP 321
DB 281 VHSQARDRPHILT-----GGHEVWRTGKFVNVPRDQDTWIFPGPTGAAVY 330
QY 322 KVDVPGDYVLVDHAFRAFNKAGLILKYGEENHEIYSHKOTDAYVLPFGA 373
DB 331 TTBQPIYAVVNHNLLEAFELGAAAFKVTGDMNDLMT-----TVRSFSGS 377

RESULT 8
NIR_ALCPA STANDARD; PRT; 376 AA.
AC P38501;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Copper-containing nitrite reductase precursor (EC 1.7.2.1) (Cu-NIR)
GN NIKR OR NIR.
OS Alcaligenes faecalis.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Alcaligenaceae; Alcaligenes.
OX NCBI_Taxid=511;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 34-48 AND 192-240.
RC STRAIN=S-6;
RX MEDLINE=93294530; PubMed=8515232;
RA Nishiyama M., Suzuki J., Kukimoto M., Ohnuki T., Horinouchi S.,
RT Beppu T.;
RT "Cloning and characterization of a nitrite reductase gene from
RL Alcaligenes faecalis and its expression in Escherichia coli.";
RL J. Gen. Microbiol. 139:725-733(1993).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS), AND MUTAGENESIS.
RC STRAIN=S-6;
RX MEDLINE=94227056; PubMed=8172899;
RA Kukimoto M., Nishiyama M., Murphy M.E.P., Turley S., Adman E.T.,
RT Horinouchi S., Beppu T.;
RT "X-ray structure and site-directed mutagenesis of a nitrite reductase
RL from Alcaligenes faecalis S-6: roles of two copper atoms in nitrite
RT reduction.";
RL Biochemistry 33:5246-5252(1994).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
RX MEDLINE=96032602; PubMed=7547950;
RA Murphy M.E., Turley S., Kukimoto M., Nishiyama M., Horinouchi S.,
RT Sasaki H., Tanokura M., Adman E.T.;
RT "Structure of Alcaligenes faecalis nitrite reductase and a copper
RL site mutant, M150E, that contains zinc.";
RL Biochemistry 34:12107-12117(1995).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
RX MEDLINE=98019216; PubMed=9353305;
RA Murphy M.E., Turley S., Adman E.T.;
RT "Structure of nitrite bound to copper-containing nitrite reductase
RL from Alcaligenes faecalis. Mechanistic implications.";
RL J. Biol. Chem. 272:28455-28460(1997).
CC -1- CATALYTIC ACTIVITY: Nitrite oxide + H(2)O + ferrioxochrome c =
CC nitrite + ferrioxochrome c.
CC -1- COFACTOR: CONTAINS ONE ATOM OF TYPE I COPPER AND ONE ATOM OF TYPE
CC II COPPER. FAD. TYPE I COPPER IS BOUND WITHIN A SINGLE MONOMER,
CC WHILE TYPE 2 COPPER IS HELD BY RESIDUES FROM EACH OF TWO MONOMERS
CC OF THE TRIMER. NITRITE IS BOUND TO THE TYPE 2 COPPER SITE.
CC PSEUDOMAZURIN IS THE PHYSIOLOGICAL ELECTRON DONOR FOR THE CU-NIR IN
CC VITRO (BY SIMILARITY).
CC -1- PATHWAY: Nitrate assimilation (denitrification).
CC -1- SUBUNIT: Homotrimer.
CC -1- SUBCELLULAR LOCATION: Periplasmic.
CC -1- INDUCTION: UNDER ANAEROBIC GROWTH CONDITIONS AND BY NITRITE.

```

```

CC -1- DOMAIN: THE TYPE I COPPER SITE IN NIR PLAYS A CRUCIAL ROLE FOR
CC ELECTRON TRANSFER FROM PSEUDOMAZURIN TO THE TYPE II COPPER SITE
CC OF NIR, WHICH COMPRISES THE CATALYTIC CENTER OF NIR FOR THE
CC REDUCTION OF NITRITE.
CC -1- SIMILARITY: Contains 2 plastocyanin-like domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D13155; BAA02440.1; -.
CC PIR: I39582; I39582.
CC PDB: 2AFN; 01-AUG-96.
CC PDB: 1NTD; 08-NOV-96.
CC PDB: 1AQ8; 25-FEB-98.
CC PDB: 1AS6; 25-FEB-98.
CC PDB: 1AS7; 25-FEB-98.
CC PDB: 1AS8; 25-FEB-98.
CC PDB: 1ERT5; 24-AUG-00.
CC PDB: 1ERT7; 24-AUG-00.
CC PDB: 1ERT8; 24-AUG-00.
CC InterPro: IPR001117; Cu-oxidase.
CC InterPro: IPR001287; CuNO2_reductase.
CC InterPro: IPR006311; Tac.
CC Pfam: PF00394; Cu-oxidase; 2.
CC PRINTS: PR00695; CUNO2RDTASE.
CC DR TIGRFAMs: TIGR01409; TAT signal_seq; 1.
CC Oxidoreductase; Copper; Metal-binding; Flavoprotein; FAD;
CC Nitrate assimilation; Repeat; Periplasmic; Signal; 3d-structure;
CC Pyruvate carboxylic acid.
CC KW
CC SIGNAL 1 33
CC CHAIN 34 376
CC DOMAIN 34 211
CC MOD RES 212 376
CC METAL 34 34
CC METAL 131 131
CC METAL 136 136
CC METAL 171 171
CC METAL 172 172
CC METAL 181 181
CC METAL 186 186
CC METAL 342 342
CC MUTAGEN 171 171
CC MUTAGEN 186 186
CC -----
CC COPPER-CONTAINING NITRITE REDUCTASE.
CC PLASTOCYANIN-LIKE 1.
CC PYRROLIDONE CARBOXYLIC ACID.
CC COPPER (TYPE 1).
CC COPPER (TYPE 2).
CC COPPER (TYPE 2).
CC COPPER (TYPE 1).
CC COPPER (TYPE 1).
CC COPPER (TYPE 1).
CC COPPER (TYPE 2).
CC H->K: LOOSES NITRITE-REDUCING ACTIVITY.
CC M->E: CONTAINS ONLY A TYPE II COPPER ATOM
CC AND FAILS TO CATALYZE THE REDUCTION OF
CC NITRITE.
CC -----
CC HELIX 42 47
CC STRAND 50 52
CC TURN 58 59
CC STRAND 67 67
CC TURN 74 74
CC STRAND 90 91
CC TURN 94 100
CC STRAND 101 102
CC TURN 108 112
CC STRAND 113 114
CC TURN 116 123
CC STRAND 125 126
CC TURN 131 131
CC STRAND 134 135
CC TURN 136 137
CC TURN 141 147
CC HELIX 141 147
CC STRAND 150 150
CC TURN 152 153
CC STRAND 154 161
CC TURN 166 171
CC STRAND 171 171
CC TURN 175 176
CC HELIX 178 182
CC TURN 183 185

```

FT STRAND 187 193  
 FT TURN 195 195  
 FT STRAND 196 196  
 FT STRAND 198 199  
 FT TURN 201 202  
 FT STRAND 205 206  
 FT STRAND 210 219  
 FT STRAND 223 223  
 FT TURN 225 226  
 FT STRAND 229 229  
 FT HELIX 235 246  
 FT TURN 247 249  
 FT STRAND 253 256  
 FT TURN 257 258  
 FT STRAND 259 259  
 FT TURN 260 263  
 FT HELIX 265 267  
 FT STRAND 269 272  
 FT TURN 273 274  
 FT STRAND 276 283  
 FT STRAND 288 288  
 FT STRAND 290 293  
 FT TURN 294 294  
 FT STRAND 297 301  
 FT TURN 302 303  
 FT STRAND 305 305  
 FT TURN 306 307  
 FT STRAND 311 314  
 FT STRAND 319 319  
 FT TURN 321 322  
 FT STRAND 323 330  
 FT STRAND 335 341  
 FT HELIX 344 348  
 FT TURN 349 349  
 FT STRAND 353 359  
 FT TURN 364 366  
 FT STRAND 367 374  
 SO SEQUENCE 376 AA; 40332 MW; 0AD918988301BP29 CRC64;

Query Match 16.3%; Score 427.5; DB 1; Length 376;  
 Best Local Similarity 33.7%; Pred. No. 8.1e-23;  
 Matches 129; Conservative 51; Mismatches 158; Indels 45; Gaps 14;

QY 15 LSLMLSGCSNQADKAAQPKSSIVDA---AAKTAANDNAASQEHCELPVIDAIYTHAP 70  
 DB 13 LAGAAALAG-----ALAPVLAITTSAMGOGAVRKATAEIALAPRO-KVELVDPPEVYHAA 64  
 QY 71 EVPEPVDRDHPAKVVMGHEVTEKVMRLAD-GVEYQFWTFGGQVPGQMRVREGDTIEVOF 129  
 DB 65 S---QVAEGGPKVVEFTWVIEBKIVIDDAGTEVHAMAFNGITVPGPLMVVHODDYLELT 121  
 QY 130 SNHPDSKMPHNVDFHAATGPGGAASFTAPGHTSTFSPALOPGLVYVHCAVAPVGM-- 187  
 DB 122 INETNTLMNIDPHATGALGGGGLTEINPGEKITLRFKATKRGVGVHNC--APPGMVP 179  
 QY 188 -HANGMYGLILVEPKGCLP-----KVDKEYVVMGDPYV---KGKXGE---QGLOR 232  
 DB 180 MHWVSGNNGAIMVLPREGLDHGKSKALTYKIIYVGGODYVVPDENGKTKKYEAPBDAY 239  
 QY 233 FDMKKAIRE-DAEVYVFNQSGVAGALTGENALKAKGETVRLFFVNGGPNLTSSFFHVGIEIF 291  
 DB 240 EDVYVQVMTLTPPHVFNQSGVAGALTGSKAMTAAGVE--KLIIVHSQANRDRPHLIGSHG 297  
 QY 292 DKVHPEGKGENHNI--QTLIPAGGAITEFKVDVPGDYVLDHAI FRAFNKGALGILK 349  
 DB 298 DYWMATKFTMPRPVDETFWTFPGGAAGAAFYTFQCGIYAVVYVHNLIEAFELGAANAFK 357  
 QY 350 VEGEENHEIYSHKQTDVAVYLPBG 372  
 DB 358 VTGEMNDLMT-----SVLAPSG 375

RESULT 9

NIR\_ALCXX  
 ID NIR\_ALCXX STANDARD; PRT; 330 AA.  
 AC P81445;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Copper-containing nitrite reductase (EC 1.7.2.1) (Cu-NIR).  
 GN NIKR.  
 OS Alcaligenes xylosoxydans (Achromobacter xylosoxydans).  
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
 OC Alcaligenaceae; Achromobacter.  
 OX NCBI TaxID=515;  
 RN [1]  
 RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS).  
 RC STRAIN=NCIMB 11015;  
 RA Dodd F.E., Haenlein S.S., Abraham Z.H.L., Eady R.R., Smith B.E.;  
 RT "Structures of a blue-copper nitrite reductase and its substrate-bound complex."  
 RL Acta Crystallogr. D 53:406-418(1997).  
 RN [2]  
 RP CHARACTERIZATION.  
 RC STRAIN=NCIMB 11015;  
 RX MEDLINE=94059002; PubMed=8240262;  
 RA Abraham Z.H.L., Lowe D.J., Smith B.E.;  
 RT "Purification and characterization of the dissimilatory nitrite reductase from Alcaligenes xylosoxydans subsp. xylosoxydans (N.C.I.M.B. 11015): evidence for the presence of both type 1 and type 2 copper centres."  
 RT Biochem. J. 295:587-593(1993).  
 RL -1- CATALYTIC ACTIVITY: Nitric oxide + H(2)O + ferricytochrome c = nitrite + ferrocyclochrome c.  
 CC -1- COPFACTOR: CONTAINS ONE ATOM OF TYPE 1 COPPER AND ONE ATOM OF TYPE II COPPER. PAD, TYPE 1 COPPER IS BOUND WITHIN A SINGLE MONOMER WHILE TYPE 2 COPPER IS HELD BY RESIDUES FROM EACH OF TWO MONOMERS OF THE TRIMER. NITRITE IS BOUND TO THE TYPE 2 COPPER SITE.  
 CC PSEUDOAZURIN IS THE PHYSIOLOGICAL ELECTRON DONOR FOR THE CU-NIR IN VITRO.  
 CC -1- PATHWAY: Nitrate assimilation (denitrification).  
 CC -1- SUBUNIT: Homotrimer.  
 CC -1- SUBCELLULAR LOCATION: Periplasmic (By similarity).  
 CC -1- DOMAIN: THE TYPE 1 COPPER SITE IN NIR PLAYS A CRUCIAL ROLE FOR ELECTRON TRANSFER FROM PSEUDOAZURIN TO THE TYPE II COPPER SITE OF NIR, WHICH COMPRISES THE CATALYTIC CENTER OF NIR FOR THE REDUCTION OF NITRITE.  
 CC -1- SIMILARITY: Contains 2 plastocyanin-like domains.  
 CC PDB; INDS; 07-JUL-97.  
 DR PDB; INDS; 07-JUL-97.  
 DR InterPro; IPR001117; Cu-oxidase.  
 DR InterPro; IPR001287; CuNO2 reductase.  
 DR Pfam; PF00394; Cu-oxidase; 2.  
 DR PRINTS; PR00695; CUNO2RDRASE.  
 KW Oxidoreductase; Copper; Metal-binding; Flavoprotein; PAD; Nitrate assimilation; Repeat; Periplasmic; 3D-structure.  
 FT DOMAIN 1 165  
 FT METAL 166 330 PLASTOCYANIN-LIKE 2.  
 FT METAL 85 85 COPPER (TYPE 1).  
 FT METAL 90 90 COPPER (TYPE 2).  
 FT METAL 125 125 COPPER (TYPE 2).  
 FT METAL 126 126 COPPER (TYPE 1).  
 FT METAL 135 135 COPPER (TYPE 1).  
 FT METAL 140 140 COPPER (TYPE 1).  
 FT METAL 296 296 COPPER (TYPE 2).  
 SO SEQUENCE 330 AA; 34413 MW; E2C83CA2CEBFCE8 CRC64;

Query Match 16.1%; Score 424; DB 1; Length 330;  
 Best Local Similarity 35.5%; Pred. No. 1.2e-22;  
 Matches 116; Conservative 36; Mismatches 119; Indels 56; Gaps 11;

QY 72 VPEPVDRDHPAKVVMGHEVTEKVMRL-ARGVEYQFWTFGGQVPGQMRVREGD 123  
 DB 10 VAPPLVHPISQVAAAGAPKVVQPFMSIEKKVMADDDGTTAAQMTFNGSPGRLVYVHGD 69  
 QY 124 TIEVQSNHPDSKMPHNVDFHAATGPGGAASFTAPGHTSTFSPALOPGLVYVHCAVA 183

```

Db      70 YIELTVNATSMFPHVDFHAATGALGGAGLQVVGQDAVIRFADRSGETFVHC--A 127
      184 PVGM--HIANGMYGLIVEPKKGLPKV-----DKERYVMQGFY-----TKKXGE-- 227
      128 PAGWVPMHVSGMNGMNLVLPDGLRDAGALAYRYVTIGESDLYVPKADGNSDIP 187
      228 -----QGIQPFDMKAIREDAEYVNGSVGALTGEMALKAKGETVRLPVG 274
      188 ALASAVADTVAVMRLTLP-----SHAVFNGAVGALGANLTAAGESVLI-- 234
      275 NGSPVLTSSFHYIGELFDKVFEGKGENH---NIOTLILPAGGAITEFKVDVEDYL 331
      235 HSQARNSRPHLIGHGDMV-WTTGKFAHPOLNMTWPIPGSAAALYTRKQPTVAY 293
      332 VDAIFRAFNKGAIGLKVEGSENHRI 358
      294 LSHNLIEMELGAAQAQASVEGQWDDDL 320

```

## RESULT 10

PCOA\_ECOLI STANDARD; PRT; 605 AA.

```

AC 04752: 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Copper resistance protein A precursor.
GN PCOA.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / ED8739;
RX MEDLINE=96130847; PubMed=8594334;
RA Brown N.L., Barrett S.R., Camakaris J., Lee B.T.O., Rouch D.A.;
RT "Molecular genetics and transport analysis of the copper-resistance
  determinant (pco) from Escherichia coli plasmid pRI1004.";
RL Mol. Microbiol. 17:1153-1166(1995).
CC - FUNCTION: REQUIRED FOR THE COPPER-INDUCIBLE EXPRESSION OF COPPER
  RESISTANCE. MAY HAVE OXIDASE ACTIVITY (BY SIMILARITY).
CC - SUBCELLULAR LOCATION: Periplasmic (Potential).
CC - SIMILARITY: BELONGS TO THE FAMILY OF MULTICOPPER OXIDASES. COPA
  SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@sib.ch).
CC -----
CC EMBL, X83541; CAA58525.1; -
CC PIR, S70159; S52253.
CC -----
CC HSSP, P36649; 1K77.
CC InterPro, IPR006376; COPA.
CC InterPro, IPR001117; Cu-oxidase.
CC InterPro, IPR002355; MultiCu_oxidase2.
CC InterPro, IPR006311; Tat.
CC Pfam, PF00394; Cu-oxidase; 2.
CC TIGRfams, TIGR01480; copper_res_A; 1.
CC TIGRfams, TIGR01409; tat_signal_seq; 1.
CC PROSITE, PS00079; MULTICOPPER_OXIDASE; 1.
CC PROSITE, PS00080; MULTICOPPER_OXIDASE2; 1.
CC Oxidoreductase; Copper; Repeat; Signal; Plasmid; Metal-binding;
KW Periplasmic.
FT SIGNAL 1 40 POTENTIAL.
FT CHAIN 41 605 COPPER RESISTANCE PROTEIN A.
FT DOMAIN 382 429 3 X 8 AA TANDEM REPEATS OF D-H-X-X-M-X-

```

```

FT REPEAT 382 389 G-M.
FT REPEAT 414 421 1.
FT REPEAT 422 429 2.
FT METAL 100 100 3.
FT METAL 102 102 COPPER (TYPE 2) (BY SIMILARITY).
FT METAL 142 142 COPPER (TYPE 3) (BY SIMILARITY).
FT METAL 144 144 COPPER (TYPE 3) (BY SIMILARITY).
FT METAL 538 538 COPPER (TYPE 1) (POTENTIAL).
FT METAL 541 541 COPPER (TYPE 2) (POTENTIAL).
FT METAL 543 543 COPPER (TYPE 3) (POTENTIAL).
FT METAL 586 586 COPPER (TYPE 3) (POTENTIAL).
FT METAL 587 587 COPPER (TYPE 1) (POTENTIAL).
FT METAL 588 588 COPPER (TYPE 1) (POTENTIAL).
FT METAL 592 592 COPPER (TYPE 3) (POTENTIAL).
FT METAL 597 597 COPPER (TYPE 1) (POTENTIAL).
SQ SEQUENCE 605 AA; 67307 MW; 8EECA182D56B27D8 CRC64;

```

Query Match Best Local Similarity 7.7%; Score 203.5; DB 1; Length 605; Matches 89; Conservative 88; Mismatches 167; Indels 83; Gaps 14;

```

QY 100 GVEYQFMFGGVPGQMTVRREGDTIEVQFSNHPDSK-----MHNVDFFHAATGCG 150
DB 60 GSERQAKTINGLPGPVLRMKRGDTITLKXNRLEQTSIHWHGILLPANMD----- 111
QY 151 GGAESAFTA--PGHTSPFKALQGLVYVHCAPVGMHIANGYGLIVEPKKGLP-X 207
DB 112 GVPGLSFGICIEPDDTYVTFKVKQNGTYWH--SHSGLOEGGVGALIIDAREPEPA 168
QY 208 VDKEYVYVQGFYTKGYGEQGLQPFDMKAIREDAEYVFN---GSVGLATGENALKA 263
DB 159 YDREHVVLMSWTDE-----NPHSLKKLKKQSDYVNFNPKPVGSPFRVNRGLSA 220
QY 264 KVGETVRLPVGNGGNLTSFHYIGELFDKVFEGKSENNHNIOTTLIPA-----CG 315
DB 221 TIAD--RQWMAEMKKNPFDLADVSG--YTYTYLNGQAPLKNMTGLFPGEKIRLRFING 276
QY 316 AATFEFKVDVGDVYLVDAIFRAFNKGAIGLKVEGSENHRIYSHKOTDAVYL----- 369
DB 277 SMATYFDIRIRGLKATVVAAGQYVNPVTVEFRFAVAETDYVIEPGEATTTAQSMD 336
QY 370 -----PEGAPQAIIDQEAQKTPAPANTLOEIQAKRATYD-----SNC 406
DB 337 RTGVARTGLATREGISAIV-----PLDRLPLITMEDMGMGCHDMAGMDHSQMGMDNS 392
QY 407 AACHPQDGKGVPN--AFPLANSDYLNADHARAASI 440
DB 393 GEMMSMDGADLPDSGTSAPMDHSSMAGMDHSSRMAGM 429

```

RESULT 11

ASO\_CUCPM STANDARD; PRT; 552 AA.

```

AC P37064;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE L-ascorbate oxidase (EC 1.10.3.3) (Ascorbate) (ASO).
OS Cucurbita pepo var. meloepo (Zucchini).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eutroside 1; Cucurbitales; Cucurbitaceae; Cucurbita.
OX NCBI_TaxID=3665;
RN [1]
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).
RX MEDLINE=92194315; PubMed=1548698;
RA Messerschmidt A., Ladenstein R., Huber R., Bolognesi M.,
  Avigliano L., Petruzzelli R., Rossi A., Finazzi-Aro A.;
  "Refined crystal structure of ascorbate oxidase at 1.9-A resolution.";
  J. Mol. Biol. 224:179-205 (1992).
RL [2]
RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).

```

RX MEDLINE=89236417; PubMed=2716059;  
RA Meserich A., Rosei A., Ladenstein R., Huber R., Bolognesi M.,  
"X-ray crystal structure of the blue oxidase ascorbate from  
RT zucchini. Analysis of the polypeptide fold and a model of the copper  
RT sites and ligands."  
RL J. Mol. Biol. 206:513-529 (1989).  
CC -1- FUNCTION: MAY BE INVOLVED IN A REDOX SYSTEM INVOLVING ASCORBIC  
CC ACID.  
CC -1- CATALYTIC ACTIVITY: 2 L-ascorbate + O(2) = 2 dehydroascorbate + 2  
CC H(2O).  
CC -1- COFACTOR: THIS PROTEIN BELONGS TO THE MULTICOPPER OXIDASES WHICH  
CC CONTAIN THREE DISTINCT CU CENTERS KNOWN AS TYPE 1 OR BLUE, TYPE  
CC 2 OR NORMAL, AND TYPE 3 OR COUPLED BINUCLEAR.  
CC -1- SIMILARITY: BELONGS TO THE FAMILY OF MULTICOPPER OXIDASES.  
CC -1- SIMILARITY: Contains 3 placocyanin-like domains.  
DR PIR: A51027; A51027.  
DR PDB: 1A0Z; 31-OCT-93.  
DR PDB: 1ASQ; 31-JAN-94.  
DR PDB: 1ASP; 31-JAN-94.  
DR PDB: 1ASQ; 31-JAN-94.  
DR GlycosuiteDB; P37064; .  
DR InterPro; IPR001117; Cu-oxidase.  
DR InterPro; IPR002355; MultiCu-oxidase2.  
DR Pfam; PF00394; Cu-oxidase; 3.  
DR PROSITE; PS00079; MULTICOPPER\_OXIDASE1; 1.  
DR PROSITE; PS00080; MULTICOPPER\_OXIDASE2; 1.  
KW Oxidoreductase; Copper; Glycoprotein; 3D-structure; Repeat.  
FT DOMAIN 1 122 PLASTOCYANIN-LIKE 1.  
FT DOMAIN 134 300 PLASTOCYANIN-LIKE 2.  
FT DOMAIN 344 523 PLASTOCYANIN-LIKE 3.  
FT DISULFID 19 201  
FT DISULFID 81 538  
FT DISULFID 180 193  
FT CARBOHYD 92 92  
FT METAL 60 60 N-LINKED (GLCNAC. . .).  
FT METAL 62 62 /FTID=CAR.000149.  
FT METAL 104 104 COPPER (TYPE 2).  
FT METAL 106 104 COPPER (TYPE 3).  
FT METAL 445 445 COPPER (TYPE 3).  
FT METAL 448 445 COPPER (TYPE 1).  
FT METAL 450 450 COPPER (TYPE 2).  
FT METAL 506 506 COPPER (TYPE 3).  
FT METAL 507 507 COPPER (TYPE 3).  
FT METAL 508 507 COPPER (TYPE 1).  
FT METAL 512 508 COPPER (TYPE 3).  
FT METAL 517 517 COPPER (TYPE 1).  
FT METAL 517 517 COPPER (TYPE 1).  
FT STRAND 3 15  
FT TURN 17 18  
FT TURN 22 27  
FT TURN 28 29  
FT STRAND 30 30  
FT STRAND 36 39  
FT TURN 40 41  
FT STRAND 43 50  
FT STRAND 57 57  
FT STRAND 59 62  
FT TURN 63 63  
FT TURN 67 68  
FT HELIX 70 72  
FT STRAND 76 76  
FT STRAND 77 79  
FT TURN 80 80  
FT STRAND 83 83  
FT TURN 85 86  
FT STRAND 88 94  
FT STRAND 99 105  
FT TURN 108 110  
FT HELIX 111 113  
FT HELIX 114 114  
FT STRAND 116 122  
FT TURN 125 126

FT STRAND 135 144  
FT HELIX 149 155  
FT TURN 156 157  
FT STRAND 169 172  
FT TURN 173 174  
FT STRAND 175 175  
FT STRAND 181 181  
FT TURN 182 183  
FT HELIX 184 186  
FT TURN 188 189  
FT TURN 192 192  
FT TURN 200 201  
FT STRAND 206 208  
FT TURN 210 211  
FT STRAND 213 220  
FT STRAND 226 231  
FT TURN 232 233  
FT TURN 236 241  
FT STRAND 242 243  
FT STRAND 244 251  
FT STRAND 254 256  
FT TURN 258 259  
FT STRAND 261 267  
FT TURN 272 273  
FT STRAND 276 283  
FT STRAND 292 298  
FT TURN 299 300  
FT TURN 303 304  
FT TURN 317 318  
FT HELIX 320 327  
FT TURN 328 329  
FT STRAND 332 332  
FT TURN 334 335  
FT STRAND 345 356  
FT TURN 357 358  
FT STRAND 359 364  
FT STRAND 365 366  
FT STRAND 367 369  
FT HELIX 376 381  
FT TURN 382 383  
FT TURN 385 386  
FT TURN 389 400  
FT TURN 403 404  
FT TURN 410 411  
FT STRAND 413 415  
FT STRAND 419 421  
FT TURN 423 424  
FT STRAND 426 433  
FT TURN 439 440  
FT STRAND 445 449  
FT TURN 450 451  
FT STRAND 454 461  
FT HELIX 466 471  
FT STRAND 479 480  
FT STRAND 482 485  
FT TURN 487 488  
FT STRAND 489 496  
FT STRAND 501 507  
FT HELIX 510 514  
FT TURN 515 516  
FT STRAND 518 523  
FT HELIX 525 527  
FT HELIX 533 536  
FT HELIX 539 545  
SQ SEQUENCE 552 AA; 61704 MW; 24660B0F47AB54B4 CRC64;

## Query Match

Best Local Similarity 6.5%; Score 170; DB 1; Length 552;  
Matches 81; Conservative 47; Mismatches 129; Indels 80; Gaps 18;

QY 101 VEQYFWT-----FGGVPGQMIRVREGDTIVQFSN--HPDSKMPHNVDFPAANG 148  
Db 10 VEQYFWAPNCNENIVMGINGQFPFGPTIRANAGDSVVELTNKLHTEGVVH--WHGIHQ 66



```

QY 149 PG-----GGAEASFTL--PGHTSTSPFKALQPLGVYHCAVAPVGMHIANMVGILVLP 201
DB 67 RGTPMADGTASISQCAINNGEFTFFNFVNDNCTGFYH---CHLGMQRASGLTGLVLP 123
QY 202 KEGLPK---VDKEYVWMOGDFYTKGKYGRQ-GLQPFDMKAIK--EDAEVYVNG----- 250
DB 124 PGKKEPFHYDEINLLSDMWHQSIHKEVGLS-----SKPIRWIGEPOTILLNGGQPD 179
QY 251 -SVGALTGEN--ALKAKVGETVRLFVNGGPNLTSSFHVIGEIFDKVHEGKGGENHNIO 307
DB 180 CSIAKYDSNLEBCKLKSESCAPYIFHVSFKTYIRI-----AS 220
QY 308 TLLIPAGAAITEFK---VDVPGDYV---LVDAIFRAFNKALGILKVEGENHEI-- 358
DB 221 TLLAALNFAIGNHQLLVVEADGNVQPFYTSIDIDYSGESYVLLITDQNPSENWASV 280
QY 359 ---SHKQTD-----AVYLPBGAPOAIDQEAAPKTPA 387
DB 281 GTRARHPNTPPGTLTLNLPNSVSK-LPTSPPTPA 316

RESULT 12
ASO_CUCMA STANDARD; PRT; 579 AA.
AC P24792; Q39539;
AT 01-MAR-1992 (Rel. 21, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE L-ascorbate oxidase precursor (EC 1.10.3.3) (Ascorbase) (ASO).
GN MAO.
OS Cucurbita maxima (Pumpkin) (Winter squash).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosid I; Cucurbitales; Cucurbitaceae; Cucurbita.
OX NCBI_Taxid=3661;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 31-48.
RC STRAIN=cv. Ebisu Nankin;
RX MEDLINE=90361033; PubMed=2143984;
RA Esaka M., Harutori T., Fujisawa K., Sakajo S., Asahi T.;
RT "Molecular cloning and nucleotide sequence of full-length cDNA for
RT ascorbate oxidase from cultured pumpkin cells.";
RL Eur. J. Biochem. 191:537-541(1990).
[2]
RP SEQUENCE FROM N.A.
RX MEDLINE=97354114; PubMed=9210335;
RA Kisu Y., Harada Y., Goto M., Esaka M.;
RT "Cloning of the pumpkin ascorbate oxidase gene and analysis of a cis-
RT acting region involved in induction by auxin.";
RL Plant Cell Physiol. 38:631-637(1997).
CC -1- FUNCTION: MAY BE INVOLVED IN A REDOX SYSTEM INVOLVING ASCORBIC
CC ACID.
CC -1- CATALYTIC ACTIVITY: 2 L-ascorbate + O(2) = 2 dehydroascorbate + 2
CC H(2)O.
CC -1- COFACTOR: THIS PROTEIN BELONGS TO THE MULTICOPPER OXIDASES WHICH
CC CONTAIN THREE DISTINCT CU CENTERS KNOWN AS TYPE 1 OR BLUE, TYPE
CC 2 OR NORML, AND TYPE 3 OR COUPLED BINUCLEAR.
CC -1- SIMILARITY: BELONGS TO THE FAMILY OF MULTICOPPER OXIDASES.
CC -1- SIMILARITY: CONTAINS 3 PLASTOCYANIN-LIKE DOMAINS.
CC CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; X55779; CAA39300.1; -
CC DR EMBL; D55677; BAA09528.1; -
CC DR HSSP; P37064; IAOZ.
CC InterPro; IPR001117; Cu-oxidase.

```

```

DR InterPro; IPR002355; MultiCu_oxidase2.
DR Pfam; PF00394; Cu-oxidase; 3.
DR PROSITE; PS00079; MULTICOPPER OXIDASE1; 1.
DR PROSITE; PS00080; MULTICOPPER OXIDASE2; 1.
KW Oxidoreductase; Copper; Metal-binding; Glycoprotein; Signal; Repeat.
FT SIGNAL 1 30
FT CHAIN 31 579
FT DOMAIN 33 152 L-ASCORBATE OXIDASE.
FT DOMAIN 164 330 PLASTOCYANIN-LIKE 1.
FT DOMAIN 374 553 PLASTOCYANIN-LIKE 2.
FT DISULFID 49 231 PLASTOCYANIN-LIKE 3.
FT DISULFID 111 568 BY SIMILARITY.
FT DISULFID 210 523 BY SIMILARITY.
FT CARBOHYD 122 122 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 355 355 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 470 470 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT METAL 90 90 COPPER (TYPE 2) (PROBABLE).
FT METAL 92 92 COPPER (TYPE 3) (PROBABLE).
FT METAL 134 134 COPPER (TYPE 3) (PROBABLE).
FT METAL 136 136 COPPER (TYPE 3) (PROBABLE).
FT METAL 475 475 COPPER (TYPE 1) (PROBABLE).
FT METAL 478 478 COPPER (TYPE 2) (PROBABLE).
FT METAL 480 480 COPPER (TYPE 3) (PROBABLE).
FT METAL 536 536 COPPER (TYPE 3) (PROBABLE).
FT METAL 537 537 COPPER (TYPE 1) (PROBABLE).
FT METAL 538 538 COPPER (TYPE 3) (PROBABLE).
FT METAL 542 542 COPPER (TYPE 1) (PROBABLE).
FT METAL 547 547 COPPER (TYPE 1) (PROBABLE).
FT CONFLICT 175 175 W -> C (IN REF. 1).
SQ SEQUENCE 579 AA; 64667 MW; 8F5AF4CB07B27689 CRC64;

```

```

Query Match 6.4%; Score 167; DB 1; Length 579;
Best Local Similarity 24.6%; Pred. No. 0.00027;
Matches 83; Conservative 45; Mismatches 129; Indels 80; Gaps 19;

```

```

QY 101 VEYQVWT-----FEGVPGQMIRVBSGDTLEVOFSN--HPDSKPHNVDFHAATG 148
DB 40 VEYMFAPDCNENIWMGNGQPFPIRAMADTVLVELINKLHTEGVYH---WHGIQ 96
QY 149 PG-----GGAEASFTL--PGHTSTSPFKALQPLGVYHCAVAPVGMHIANMVGILVLP 201
DB 97 RGTPMADGTASISQCAINNGEFTFFNFVNDNCTGFYH---CHLGMQRASGLTGLVLP 153
QY 202 KEGLPK---VDKEYVWMOGDFYTKGKYGRQ-GLQPFDMKAIK--EDAEVYVNG----- 250
DB 154 PGKKEPFHYDEINLLSDMWHQSIHKEVGLS-----SKPIRWIGEPOTILLNGGQPD 209
QY 251 -SVGALTGEN--ALKAKVGETVRLFVNGGPNLTSSFHVIGEIFDKVHEGKGGENHNIO 307
DB 210 CSIAKYDSNLEBCKLKSESCAPYIFHVSFKTYIRI-----AS 250
QY 308 TLLIPAGAAITE---FKVDVPGDYV---LVDAIFRAFNKALGILKVEGENHEI-- 359
DB 251 TLLAALNFAIGNHQLLVVEADGNVQPFYTSIDIDYSGESYVLLITDQNPSENWASV 310
QY 360 ---SHKQTD-----AVYLPBGAPOAIDQEAAPKTPA 387
DB 311 GTRARHPNTPPGTLTLNLPNSVSK-LPTSPPTPA 346

```

```

RESULT 13
CPAL_PSESM STANDARD; PRT; 609 AA.
AC P12374;
AT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Copper resistance protein A precursor.
GN COPA.
OS Pseudomonas syringae (pv. tomatos).
OG Plasmid pPT23D.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.

```

```

OY      110 GOVVGOMIRVREGGLTIEVQFSNHPDSK-----MHNVDFFHAATPGGGAASFTA-   159
       |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB      70 GGLBGPLLRMRKEGGDTVLRLRNRRLDAATSITHMGILLPRMD-----GVPLSLFAGI   121
OY      160 -PGRHTSFSEFKALQDGLVVYHCAVAPEVMHIANGMYGLLVEPKGLP-KVDKEYVMQG   217
       |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB      122 EPGGVVUYEKFVKQONGTYWYH---SHSGFOEQGVGVPLVIEAKEPERPFKYDSEHVMLT   178
OY      218 DFLTKKGKYGEGOLQFPFMERKAIEDAEYVFNGSVGLTGEMNLKAKVGETVRLFVONG   277
DB      179 DWTFDE-----DPVSILMTLKKSQSYNYNFH-----KRIVGVGFVNDVADKW   218
OY      278 P-----NLTSEFYHGELFDKHFEPGGKGENNIQTLLIPA-----GG   315
DB      219 AATVADSRKMAEMKNPTDLADVSGATY--TYLLNGCAPRMNTGFLPREBKRLRRING   276
OY      316 AATTEPKVDVPGDYLVLDNAIFRAFNGALGIKVEGEENHEIYSHKOTDAVYLPEGAPQ   375
DB      277 SAMTYFPIRIPLGLKMVTVAADGGQVNVVEDELRIAAVEFDVIVEPTAEAYTV---FAQ   333
OY      376 AID-----TOEAPKTAPANLOEIQAIGATY-----DSCN   406
DB      334 SMORTGYARGTLAVREGLVAQVPELPDRPLVTMDMGMMGDHSGMDMSGMSGADGM   393
OY      407 AACGQPDGKGVPAFPPLANSVDLYLNADHABAASIIVANGLSGI-----TVNGN---QYE   457
DB      394 QTMSSWGSDSIP-LMDHSKRSITMOGMDHGMSCMDHGMGMGMVQSHPSABENDNPVDMQ   452
OY      458 SYMPAIALSDQIA----*-NVITY   476
DB      453 AMSPTALNDPGLGRNNRKRVITY   477

RESULT_14
CYCA_GLUOX STANDARD; PRY; 478 AA.
AC      047945; 008083;
DT      01-NOV-1997 (Rel. 35, Created)
DT      01-NOV-1997 (Rel. 35, Last sequence update)
DT      28-FEB-2003 (Rel. 41, Last annotation update)
DE      Alcohol dehydrogenase cytochrome c subunit precursor (Cytochrome c-
DE      553) (C553) (G3-ADH subunit II).
GN      ADHB.
OS      Glucobacter oxydans (Glucobacter suboxydans).
OC      Bacteria; Proteobacteria; Alphaproteobacteria; Rhodospirillales;
OC      Acetobacteraceae; Glucobacter.
CX      NCBI_TaxId=442;
RN      [1]
RP      SEQUENCE FROM N.A.
RA      Takeda Y., Shimizu T.;
RT      "Cloning and sequencing of the gene encoding cytochrome c-553 (CO)
RT      from Glucobacter suboxydans.";
RL      J. Ferment. Bioeng. 72:1-6(1991).
RL      [2]
RP      SEQUENCE FROM N.A., AND SEQUENCE OF 38-52.
RC      STRAIN=IFO 12528;
RX      MEDLINE=97208225; PubMed=9055427;
RA      Kondo K., Horinouchi S.;
RT      "Characterization of the genes encoding the three-component membrane-
RT      bound alcohol dehydrogenase from Glucobacter suboxydans and their
RT      expression in Acetobacter pasteurianus.";
RL      Appl. Environ. Microbiol. 63:1131-1138(1997).
CC      -1- FUNCTION: DEHYDRATION OF PRIMARY ALCOHOLS (EXCEPTION: METHANOL).
CC      -1- SUBUNIT: HETERODIMER (DEHYDROGENASE, CYTOCHROME AND PROTEIN
CC      ADHS), THAT FORMS THE ALCOHOL DEHYDROGENASE COMPLEX.
CC      -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND, FACING THE PERIPLASMIC
CC      SPACE (POTENTIAL).
CC      -1- PM: BINDS THREE HEME GROUPS PER MOLECULE (POTENTIAL).
CC      -----
CC      THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation at
CC      the European Bioinformatics Institute. There are no restrictions on its
```

CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; M58760; AAA4935.1; -  
DR EMBL; D86375; BAA19754.1; -  
DR HSSP; P04164; 1D11.  
DR InterPro; IPR003088; Cyt\_C1.  
DR InterPro; IPR002329; Cyt\_C1C.  
DR InterPro; IPR003219; Cyt\_C adh.  
DR InterPro; IPR000345; CytC\_heme\_bind.  
DR Pfam; PF00034; cytochrome\_c; 1.  
DR PRINTS; PR00605; CYTOCHROME C.  
DR ProDom; PD011584; CytC adh; 1.  
DR PROSITE; PS00190; CYTOCHROME C; 3.  
KW Electron transport; Oxidoreductase; Signal; Heme; Membrane;  
KW Periplasmic; Pyridinolone carboxylic acid.  
FT SIGNAL 1 36  
FT CHAIN 37 478  
FT FT ALCOHOL DEHYDROGENASE CYTOCHROME C  
FT FT SUBUNIT.  
FT MOD RES 37 37 PYRROLIDONE CARBOXYLIC ACID.  
FT BINDING 56 56 HEME 1 (COVALENT) (BY SIMILARITY).  
FT BINDING 59 59 HEME 1 (COVALENT) (BY SIMILARITY).  
FT METAL 60 60 IRON (HEME 1 AXIAL LIGAND)  
FT METAL 60 60 IRON (BY SIMILARITY).  
FT BINDING 204 204 HEME 2 (COVALENT) (BY SIMILARITY).  
FT BINDING 207 207 HEME 2 (COVALENT) (BY SIMILARITY).  
FT METAL 208 208 IRON (HEME 2 AXIAL LIGAND)  
FT METAL 208 208 IRON (BY SIMILARITY).  
FT BINDING 340 340 HEME 3 (COVALENT) (BY SIMILARITY).  
FT BINDING 343 343 HEME 3 (COVALENT) (BY SIMILARITY).  
FT METAL 344 344 IRON (HEME 3 AXIAL LIGAND)  
FT METAL 344 344 IRON (BY SIMILARITY).  
SQ SEQUENCE 478 AA; 51198 MW; 3FBC1F935AD1D866 CRC64;

Query Match 6.2%; Score 163; DB 1; Length 478;  
Best Local Similarity 36.0%; Pred. No. 0.00041;  
Matches 49; Conservative 18; Mismatches 59; Indels 10; Gaps 6;

QY 369 LPEGAPQADITDEAPRTAPALQEQIKAGKATYDSCNCAHQPPQKGYPNAPPLA-NS 427  
DB 306 VEEGKNLG---ODDQGTALLNKGGGNGAGAEVYLNCALCHMNDGTGNRWFPLLAGNP 362  
QY 428 DYLNADHRAASIVANGSGKITVNGQYESV-MPAIA--LSDQOIANYITLINSFGK 484  
DB 363 VITTDPTSLANVAVFG--GILPPTNSAPSAVAMPGFNHLSQEMADVNFRRKGWNN 420  
QY 485 G-GQSSADDAVAKAKT 499  
DB 421 APTGVSASDIQKLRFT 436

RESULT 15  
ASO\_TOBAC STANDARD; PRT; 578 AA.  
ID ASO\_TOBAC  
AC Q40588;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 15-UL-1999 (Rel. 38, Last annotation update)  
DE L-ascorbate oxidase precursor (EC 1.10.3.3) (Ascorbase) (ASO).  
GN AAO.  
OS Nicotiana tabacum (Common tobacco).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
OC Asteridae; lamiales; Solanales; Solanaceae; Nicotiana.  
OX NCBI\_TaxID=4097;  
RN [1]  
RX SEQUENCE FROM N.A.  
RP MEDLINE=96194464; PubMed=8624413;  
RA Kato N., Esaka M.,  
RT "cDNA cloning and gene expression of ascorbate oxidase in tobacco";  
RL Plant. Mol. Biol. 30:833-837(1996).

CC -1- FUNCTION: MAY BE INVOLVED IN A REDOX SYSTEM INVOLVING ASCORBIC  
CC ACID.  
CC -1- CATALYTIC ACTIVITY: 2 L-ascorbate + O(2) = 2 dehydroascorbate + 2  
CC H(2)O.  
CC -1- COPFACTOR: THIS PROTEIN BELONGS TO THE MULTICOPPER OXIDASES WHICH  
CC CONTAIN THREE DISTINCT CU CENTERS KNOWN AS TYPE 1 OR BLUE, TYPE  
CC 2 OR NORMAL, AND TYPE 3 OR COUPLED BINUCLEAR.  
CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN YOUNG AND GROWING TISSUES.  
CC -1- SIMILARITY: BELONGS TO THE FAMILY OF MULTICOPPER OXIDASES.  
CC -1- SIMILARITY: Contains 3 plastocyanin-like domains.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; D43624; BAA07734.1; -  
DR PIR; S66353; S66353.  
DR HSSP; P37064; 1A02.  
DR InterPro; IPR001117; Cu-oxidase.  
DR InterPro; IPR002355; Multicopper oxidase2.  
DR Pfam; PF00394; Cu-oxidase; 3.  
DR PROSITE; PS00079; MULTICOPPER OXIDASE1; 1.  
DR PROSITE; PS00080; MULTICOPPER OXIDASE2; 1.  
KW Signal; Oxidoreductase; Copper; Glycoprotein; Repeat.  
FT SIGNAL 1 28  
FT CHAIN 29 578  
FT DOMAIN 30 149  
FT DOMAIN 161 328  
FT DOMAIN 372 550  
FT DISULFID 108 565  
FT DISULFID 207 221  
FT METAL 87 87  
FT METAL 89 89  
FT METAL 131 131  
FT METAL 133 133  
FT METAL 472 472  
FT METAL 475 475  
FT METAL 477 477  
FT METAL 533 533  
FT METAL 534 534  
FT METAL 535 535  
FT METAL 539 539  
FT METAL 544 544  
FT CARBOHYD 206 206  
FT CARBOHYD 349 349  
FT CARBOHYD 394 394  
FT CARBOHYD 438 438  
FT CARBOHYD 451 451  
SQ SEQUENCE 578 AA; 64865 MW; C2371F5FEA353E3B CRC64;

Query Match 6.2%; Score 162; DB 1; Length 578;  
Best Local Similarity 22.9%; Pred. No. 0.00061;  
Matches 104; Conservative 55; Mismatches 162; Indels 134; Gaps 23;

QY 91 VEKVRNLADGVYQFWTEGGQVPGQMIIVREEDTTEVQFSN--HPDSKMPHNVDPHAATG 148  
DB 37 VEYIHWSPDGEESVVMGJNGQFPPTIRAKAGDTVAVHLTNKLTGEGVYH---WGIHQ 93  
QY 149 PG-----GGAASFLPA--PGHTSTSPKALQPLGVNCAVAPVGMHLANGVYGLIVER 201  
DB 94 IGTTPADGTALISQCAINPGETFLRFYVDKAGTYFYH--GHYMQSAGLYGSLIVEV 150  
QY 202 KEGLPK---VDKEYVWGDDFTYTKGYEGOGLOPFDMEKALREDAVYVFNFGVATLGE 258  
DB 151 GEGEKEPFFHYDGEFPLLLSDWMHKGSH-----QEVDL-----SS 185  
QY 259 NALKAKVGETVRLVY-GNGGPNLTSSFVIGEIFDKVHEEGKGNNHI----- 306  
DB 186 NPLR-WIEGPOTLLNGRGQYVNCISLARFSKPLPQCKLRGGEQVAPQILRVPRNKIYRL 244

```

QY 307 ---QTLIPAGCAITEFK---VDVPGDYLVDAIFRAFNKGALGLKVEGENHEIYS 360
DB 245 RAASTALSLSLAIGCHKVNVLEADGNV-----QPS-----VQMDIYS 286
QY 361 HKQTDVYLPEGAPO-----AIDTO-EAPKTPANLQEOIKAKATYDSCAACHQPDG 414
DB 287 GESYSVLFKTDQPTKNWYSINVRGEPKTPGGLT-----LNYLPSNAS 332
QY 415 KGVNAPFPLAN--SDYLNADHRAAASIVANGLSGK-----ITVNG----- 453
DB 333 -KFTLPPLPIAPLMNDY---NHSKSPSNKIFALMGSPKPPQNRRIILNTONKIDGYT 388
QY 454 ---NGYESVMPAIALSDQIANVITYTLNSFGNK 484
DB 389 KMAINNVSLVLPF-----QYLGSIKYGIAFDTK 418

RESULT 16
C552_THETH STANDARD; PRT; 148 AA.
ID P04164;
AC 20-MAR-1987 (Rel. 04, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Cytochrome c-552 precursor (C552).
GN CYCA.
OS Thermus thermophilus.
OC Bacteria; Deinococcus-Thermus; Deinococci; Thermales; Thermaceae;
OC Thermus.
OX NCBI_TaxId=274;
RN [1]
RA SEQUENCE FROM N.A.
RA Keightley J.A., Wachter M.W., Fee J.A.;
RA Submitted (MAY-1992) to the EMBL/GenBank/DBJ databases.
RN [2]
RA SEQUENCE OF 18-148.
RA STRAIN=HB8 / ATCC 27634;
RA MEDLINE=8519131; PubMed=2986625;
RA Titani K., Ericsson L.H., Hon-Nami K., Miyazawa T.;
RA "Amino acid sequence of cytochrome c-552 from Thermus thermophilus
RA HB8."
RA Biochem. Biophys. Res. Commun. 128:781-787(1985).
RN [3]
RA X-RAY CRYSTALLOGRAPHY (1.28 ANGSTROMS).
RA MEDLINE=97428333; PubMed=9281430;
RA Than M.E., Hof P., Huber R., Bouterkov G.P., Bartunik H.D., Buse G.,
RA Soulimane T.;
RA "Thermus thermophilus cytochrome-c552: a new highly thermostable
RA cytochrome-c structure obtained by MAD phasing."
RA J. Mol. Biol. 271:629-644(1997).
RN [4]
RA X-RAY CRYSTALLOGRAPHY (1.80 ANGSTROMS) OF 20-148.
RA STRAIN=HB8 / ATCC 27634;
RA MEDLINE=21025441; PubMed=11152119;
RA Fee J.A., Chen Y., Todaro T.R., Bren K.L., Patel K.M., Hill M.G.,
RA Gomez-Moran E., Loshr T.M., Ai J., Thony-Meyer L., Williams P.A.,
RA Stura E., Stridhar V., McKee D.E.;
RA "Integrity of Thermus thermophilus cytochrome c552 synthesized by
RA Escherichia coli cells expressing the host-specific cytochrome c
RA maturation genes, ccmABCDGHI: biochemical, spectral, and structural
RA characterization of the recombinant protein."
RA Protein Sci. 9:2074-2084(2000).
RN [5]
RA FUNCTION: THIS MONOMER BASIC PROTEIN APPEARS TO FUNCTION AS AN
RA ELECTRON DONOR TO CYTOCHROME OXIDASE IN T.THERMOPHILUS. IT MAY
RA BE A C2-TYPE CYTOCHROME BUT IS DISTANTLY RELATED TO OTHER
RA CYTOCHROMES.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

```

CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
DB EMBL: M93437; AAB8580.1; -.
DR PDB: 1C52; 24-JUN-98.
DR PDB: 1DT1; 28-MAR-01.
DR PDB: 1FOC; 09-OCT-02.
DR InterPro: IPR003088; Cyt C1.
DR InterPro: IPR003045; CytC_heme_bind.
DR Pfam: PF00034; cytochrome_c_1.
DR ProSite: PS00190; CYTOCHROME_C_1.
KW Electron transport; Heme; Signal; 3D-structure.
FT CHAIN 1 18 CYTOCHROME C-552.
FT BINDING 28 28 HEME (COVALENT).
FT BINDING 31 31 HEME (COVALENT).
FT METAL 32 32 IRON (HEME AXIAL LIGAND).
FT METAL 80 80 IRON (HEME AXIAL LIGAND) (OR M-86).
FT HELIX 21 24
FT TURN 25 25
FT HELIX 26 32
FT TURN 34 35
FT STRAND 39 39
FT TURN 40 42
FT STRAND 43 43
FT TURN 47 48
FT HELIX 49 54
FT TURN 55 55
FT TURN 57 58
FT HELIX 59 69
FT STRAND 71 77
FT TURN 78 79
FT STRAND 80 86
FT TURN 90 91
FT HELIX 94 106
FT TURN 107 108
FT TURN 110 110
FT HELIX 111 113
FT TURN 115 116
FT HELIX 122 128
FT TURN 129 130
FT HELIX 135 143
FT TURN 144 145
SQ SEQUENCE 148 AA; 15974 MW; 6AE7E806BDB5EALD CRC64;

Query Match 6.1%; Score 161; DB 1; Length 148;
Best Local Similarity 37.3%; Pred. No. 0.00012;
Matches 44; Conservative 23; Mismatches 39; Indels 12; Gaps 6;

QY 396 KAGKATYDSCNCAACHQPDGKGVNAPFPLAN--SDYLNADHRAAASIVA--NGLSGKITV 451
DB 18 QADGAKIVNOCAGCIGNQCGIIPGAFPLAGHVAELIAEGREYIILVLGLGQCIIV 77
QY 452 NGYESVMPAIALSDQIANVITYTLNSFGN---KGGQ-LSADDAVK--AKTKP 501
DB 78 KGMKIVGVVSSPAQKDEBIAVLNHIATAMGDAAKVKGFPTAEVKKLPAKKLTP 135

RESULT 17
ID CPA2_PSESIM STANDARD; PRT; 589 AA.
AC P59571;
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Copper resistance protein A homolog precursor.
GN COPA OR PSP03914.
OS Pseudomonas syringae (pv. tomato).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxId=323;
RN [1]
RA SEQUENCE FROM N.A.

```





CC -1- FUNCTION: MAY BE INVOLVED IN A REDOX SYSTEM INVOLVING ASCORBIC  
 CC ACID.  
 CC -1- CATALYTIC ACTIVITY: 2 L-ascorbate + O(2) = 2 dehydroascorbate + 2  
 CC H(2)O.  
 CC -1- COFACTOR: THIS PROTEIN BELONGS TO THE MULTICOPPER OXIDASES WHICH  
 CC CONTAIN THREE DISTINCT CU CENTERS KNOWN AS TYPE 1 OR BLUE, TYPE  
 CC 2 OR NORMAL, AND TYPE 3 OR COUPLED BINUCLEAR.  
 CC -1- SIMILARITY: BELONGS TO THE FAMILY OF MULTICOPPER OXIDASES.  
 CC -1- SIMILARITY: TO CERULOPLASMIN AND LACCASE.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: X64257; CAA45554.1; -.  
 CC PIR: S23763; S23763.  
 CC DR InterPro: IPR001117; Cu-oxidase.  
 CC Pfam: PF00394; Cu-oxidase; 1.  
 CC KM Oxidoreductase; Copper; Glycoprotein; Multigene family; Signal.  
 CC FT SIGNAL 1 23  
 CC FT CHAIN 24 535  
 CC FT METAL 444 444  
 CC FT CARBOHYD 33 33  
 CC FT CARBOHYD 61 61  
 CC FT CARBOHYD 110 110  
 CC FT CARBOHYD 330 330  
 CC FT CARBOHYD 350 350  
 CC FT CARBOHYD 422 422  
 CC SQ SEQUENCE 555 AA; 62130 MW; 5BP23C9D73EA6CA CRC64;  
 CC  
 CC Query Match 5.5%; Score 145; DB 1; Length 555;  
 CC Best Local Similarity 26.1%; Pred. No. 0.0091;  
 CC Matches 71; Conservative 37; Mismatches 132; Indels 32; Gaps 12;  
 CC  
 CC QY 100 GVEYQFMTGCGPGVGMIRREGDTLEVOFSNHPD-----SKMHNVDFAHATGPGG 152  
 CC DB 43 GVPQVILLNGEPFGNINSTNNVIVFNLDPEFLITNNGIGHKNCQMDGFRPGIM 102  
 CC QY 153 ABASTFAPGHTSTFSFKAL-QPGLVYVHCANAVGKMHANGLVLPKELP---K 207  
 CC DB 103 CP---IMPSTNYTHFPQPKDQIGSYFY---PTTGKMHRAAGVGGRVNSRLIIPVPYAD 156  
 CC QY 208 VKKEYVVMGDFPTKCKYGEQGLQRP-DMEKAI-REDAEVVFNQSVGLTGENA--LKA 263  
 CC DB 157 PEDDYTVLLIGDWYTK--SHTQLKPKLDDGRTIGRPDG--IYNGKSGKGDGSDALEFTL 211  
 CC QY 264 KYGETVRLFVNGGPGPLTSSFVIGEIFPKVHEFGGKGENHNIOITLIPAG--GAITE 320  
 CC DB 212 KKGKTYRVAICNVGVTSTINFRIOHKKMLVENEGSHVLONDIIDLAVHVGCGFTIYVA 271  
 CC QY 321 FRYDVPGDVLVDHALFRAFNKALGILKEVG 352  
 CC DB 272 NQ--EPKDYVVAASRFLKTVITTTGLRLYEG 301  
 CC  
 CC RESULT 21  
 CC FAS\_BOVIN STANDARD; PRT; 2211 AA.  
 CC ID AC Q28107; Q28108;  
 CC DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 CC DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 CC DE Coagulation factor V precursor (Activated protein C cofactor) .  
 CC GN F5.  
 CC OS Bos taurus (Bovine) .  
 CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 CC OC Bovidae; Bovinae; Bos.  
 CC NCBI\_Taxid=9913;  
 CC OK

RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Liver;  
 RX MEDLINE=92147638; PubMed=1737753;  
 RA Quinto E.R., Besson C.T., Mann K.G., Macgillivray R.T.;  
 RT "The complete cDNA sequence of bovine coagulation factor V.";  
 RL J. Biol. Chem. 267:2971-2978(1992).  
 CC -1- FUNCTION: COAGULATION FACTOR V IS A COFACTOR THAT PARTICIPATES  
 CC WITH FACTOR XA TO ACTIVATE PROTHROMBIN TO THROMBIN.  
 CC -1- SUBUNIT: Factor Va is composed of a heavy chain and a light  
 CC chain, noncovalently bound. The interaction between the two chains  
 CC is calcium-dependent.  
 CC -1- DOMAIN: Domain B contains 29.5 X 9 AA tandem repeats, and 2 X 14  
 CC AA repeats.  
 CC -1- PTM: Thrombin activates factor V proteolytically to the active  
 CC cofactor, factor Va (formation of a heavy chain at the N-  
 CC terminus and a light chain at the C-terminus).  
 CC -1- PTM: SULFATION IS REQUIRED FOR EFFICIENT THROMBIN CLEAVAGE AND  
 CC ACTIVATION AND FOR FULL PROCOAGULANT ACTIVITY (BY SIMILARITY).  
 CC -1- SIMILARITY: Contains 3 F5/8 type A domains.  
 CC -1- SIMILARITY: Contains 2 F5/8 type C domains.  
 CC -1- SIMILARITY: STRONG, TO COAGULATION FACTOR VIII.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: M81440; AAA30512.1; -.  
 CC DR EMBL: M81441; AAA30513.1; -.  
 CC DR PIR: A42580; KPBO5.  
 CC DR HSSP: P12259; 1CZ7.  
 CC DR InterPro: IPR001117; Cu-oxidase.  
 CC DR InterPro: IPR000421; FA58\_C.  
 CC DR Pfam: PF00394; Cu-oxidase; 3.  
 CC DR Pfam: PF00754; F5\_F8\_type\_C; 2.  
 CC DR SMART: SM00231; FA58C\_2.  
 CC DR PROSITE: PS01285; FA58C\_1; 2.  
 CC DR PROSITE: PS01286; FA58C\_2; 2.  
 CC DR PROSITE: PS50022; FA58C\_3; 2.  
 CC DR PROSITE: PS00079; MULTICOPPER OXIDASE1; 2.  
 CC KM Blood coagulation; Glycoprotein; Sulfation; Calcium; Signal; Zymogen;  
 CC Repeat.  
 CC FT SIGNAL 1 28  
 CC FT CHAIN 29 2211  
 CC FT CHAIN 29 741  
 CC FT PEPTIDE 742 1564  
 CC FT  
 CC FT CHAIN 1565 2211  
 CC FT DOMAIN 30 327  
 CC FT DOMAIN 30 193  
 CC FT DOMAIN 203 327  
 CC FT DOMAIN 348 686  
 CC FT DOMAIN 348 525  
 CC FT DOMAIN 535 686  
 CC FT DOMAIN 636 1564  
 CC FT DOMAIN 1124 1151  
 CC FT DOMAIN 1124 1137  
 CC FT REPEAT 1138 1151  
 CC FT REPEAT 1188 1453  
 CC FT  
 CC FT REPEAT 1188 1196  
 CC FT REPEAT 1197 1205  
 CC FT REPEAT 1206 1214  
 CC FT REPEAT 1215 1223  
 CC FT REPEAT 1224 1232  
 CC FT REPEAT 1233 1241  
 CC FT REPEAT 1242 1250  
 CC FT REPEAT 1251 1259  
 CC FT REPEAT 1260 1268  
 CC  
 CC POTENTIAL.  
 CC COAGULATION FACTOR V.  
 CC HEAVY CHAIN (BY SIMILARITY).  
 CC ACTIVATION PEPTIDE (CONNECTING REGION)  
 CC (BY SIMILARITY).  
 CC LIGHT CHAIN (BY SIMILARITY).  
 CC F5/8 TYPE A 1.  
 CC PLASTOCYANIN-LIKE 1.  
 CC PLASTOCYANIN-LIKE 2.  
 CC F5/8 TYPE A 2.  
 CC PLASTOCYANIN-LIKE 3.  
 CC PLASTOCYANIN-LIKE 4.  
 CC B.  
 CC 2 X 14 AA TANDEM REPEATS.  
 CC 1-1.  
 CC 1-2.  
 CC 30 X 9 AA APPROXIMATE TANDEM REPEATS OF  
 CC [AS]-L-S-P-D-[LP]-(GS)-Q-(TE).  
 CC 2-1.  
 CC 2-1.  
 CC 2-2.  
 CC 2-3.  
 CC 2-4.  
 CC 2-5.  
 CC 2-6.  
 CC 2-7.  
 CC 2-8.  
 CC 2-9.

FT	REPEAT	1269	1277	2-10.
FT	REPEAT	1278	1286	2-11.
FT	REPEAT	1287	1295	2-12.
FT	REPEAT	1296	1304	2-13.
FT	REPEAT	1305	1313	2-14.
FT	REPEAT	1314	1322	2-15.
FT	REPEAT	1323	1331	2-16.
FT	REPEAT	1332	1340	2-17.
FT	REPEAT	1341	1349	2-18.
FT	REPEAT	1350	1358	2-19.
FT	REPEAT	1359	1367	2-20.
FT	REPEAT	1368	1376	2-21.
FT	REPEAT	1377	1385	2-22.
FT	REPEAT	1386	1394	2-23.
FT	REPEAT	1395	1403	2-24.
FT	REPEAT	1404	1412	2-25.
FT	REPEAT	1413	1421	2-26.
FT	REPEAT	1422	1430	2-27.
FT	REPEAT	1431	1439	2-28.
FT	REPEAT	1440	1444	2-29 (PARTIAL).
FT	REPEAT	1445	1453	2-30.
FT	DOMAIN	1569	1590	PS/8 TYPE A 3.
FT	DOMAIN	1569	1738	PLASTOCYANIN-LIKE 5.
FT	DOMAIN	1748	1890	PLASTOCYANIN-LIKE 6.
FT	DOMAIN	1894	2048	PS/8 TYPE C 1.
FT	DOMAIN	2053	2208	PS/8 TYPE C 2.
FT	SITE	741	742	CLEAVAGE (BY THROMBIN)
FT	SITE	1034	1035	CLEAVAGE (BY THROMBIN)
FT	SITE	1564	1565	CLEAVAGE (BY THROMBIN)
FT	DISULFID	167	193	PROBABLE.
FT	DISULFID	499	525	PROBABLE.
FT	DISULFID	1712	1738	PROBABLE.
FT	DISULFID	1894	2048	PROBABLE.
FT	DISULFID	2053	2208	PROBABLE.
FT	MOD_RES	697	697	BY SIMILARITY.
FT	MOD_RES	701	701	SULFATION (POTENTIAL).
FT	MOD_RES	730	730	SULFATION (POTENTIAL).
FT	MOD_RES	1513	1513	SULFATION (POTENTIAL).
FT	MOD_RES	1529	1529	SULFATION (POTENTIAL).
FT	MOD_RES	1537	1537	SULFATION (POTENTIAL).
FT	MOD_RES	1541	1541	SULFATION (POTENTIAL).
FT	CARBOHYD	225	225	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	239	239	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	297	297	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	382	382	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	460	460	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	553	553	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	587	587	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	745	745	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	756	756	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	774	774	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	780	780	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	902	902	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	952	952	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	964	964	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	1044	1044	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	1053	1053	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	1062	1062	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	1071	1071	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	1078	1078	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	1094	1094	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	1451	1451	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	1490	1490	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	1550	1550	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	1690	1690	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	1839	1839	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	1997	1997	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	2196	2196	N-LINKED (GLCNAC. . .)
FT	VARIANT	587	592	NFTLPA -> T (IN VARIANT 2).
SO	SEQUENCE	2211	AA;	248981 MM; CBSF90B738667C45 CRC64;

Query Match Best Local Similarity 5.5%; Score 143.5; DB 1; Length 2211; Pred. No. 0.07;

Matches	80;	Conservative	50;	Mismatches	133;	Indels	121;	Gaps	16;
QY	25	NOADKAAPKSTVAAAKTANADNAAGHGGELPVIDATVTHAPEVPPDR-----	78						
DB	328	NOAKTRNPKLTDORRHIKMEYFIABE-----EVIDVAFIIPANNDKXYSRH	379						
QY	79	-----DHPKVVVXMEVEKVMRLADVEYQFMFP-----GGQVPCQMIRV	119						
DB	380	LNFPSNRIGKHKKVYK-----QYDDSFTRKLEDPSSGGDILGPPIRA	425						
QY	120	REGDTIEVQFSNHPD---SKRPHANVF-----HAATPGGAGALSFAPHTSTFSRK	169						
DB	426	QVRDTLKIYFKXMASRSYSIYPHGVTFSPYDNEVSSSTSGSNTMIRAVRPGETTYTKWN	485						
QY	170	AL---OPGLVYVHCAVAP-----VGMHANGMGYGLIYVEKPGPK-----VDKEYVV	214						
DB	486	ILIESDEPTENDAOCLTRPYYSNVDTITRDILASGLIGLICKRSISDRGICQAAIDBOA	545						
QY	215	MQGDV-YTKGKYGEQLOPF--DMEKAIREDAY-----VFNQSVGALTGENALAKVGE	267						
DB	546	VPAVFDENKSWYIEDNIYKFCENPEKVRKDDPKFESNIMSNFTLPALNG-----	595						
QY	268	TYRLVNGNGENVLTSSFHYGIEIFD-----KAFPEGCK---GENHNQ	307						
DB	596	-----YVPESIPILGFCFDDTVQWHPCSVGTQNDILTIHTGHSFIYGRHEDT	644						
QY	308	TLTLPAGGAIREFKVDVGDVYL	331						
DB	645	LTLFPGQESVT-VTMNVGTMWL	667						
RESULT 22									
QY	16	OCT-2001 (Rel. 40, Created)							
DB	16	OCT-2001 (Rel. 40, Last sequence update)							
QY	15	SEP-2003 (Rel. 42, Last annotation update)							
DB	15	SEP-2003 (Rel. 42, Last annotation update)							
QY	15	Coagulation factor V precursor (Activated protein C cofactor).							
DB	15	Coagulation factor V precursor (Activated protein C cofactor).							
QY	15	Sus scrofa (Pig).							
DB	15	Sus scrofa (Pig).							
QY	15	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;							
DB	15	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;							
QY	15	Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.							
DB	15	Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.							
QY	15	NCBI_TaxID=9823;							
DB	15	NCBI_TaxID=9823;							
QY	15	SEQUENCE FROM N.A., AND 3D-STRUCTURE MODELING OF PS/8 TYPE A AND C							
DB	15	SEQUENCE FROM N.A., AND 3D-STRUCTURE MODELING OF PS/8 TYPE A AND C							
QY	15	TISSUE=Liver;							
DB	15	TISSUE=Liver;							
QY	15	MEDLINE=21121490; PubMed=11229814;							
DB	15	MEDLINE=21121490; PubMed=11229814;							
QY	15	Kim H.K.W.;							
DB	15	Kim H.K.W.;							
QY	15	"Porcine factor V: cDNA cloning, gene mapping, three-dimensional							
DB	15	"Porcine factor V: cDNA cloning, gene mapping, three-dimensional							
QY	15	protein modeling of membrane binding sites and comparative anatomy of							
DB	15	protein modeling of membrane binding sites and comparative anatomy of							
QY	15	domains."/;							
DB	15	domains."/;							
QY	15	Cell. Mol. Life Sci. 58:148-159 (2001).							
DB	15	Cell. Mol. Life Sci. 58:148-159 (2001).							
QY	15	-1- FUNCTION: Coagulation factor V is a cofactor that participates							
DB	15	-1- FUNCTION: Coagulation factor V is a cofactor that participates							
QY	15	with factor Xa to activate prothrombin to thrombin.							
DB	15	with factor Xa to activate prothrombin to thrombin.							
QY	15	-1- SUBUNIT: Factor Va is composed of a heavy chain and a light							
DB	15	-1- SUBUNIT: Factor Va is composed of a heavy chain and a light							
QY	15	chain, noncovalently bound. The interaction between the two chains							
DB	15	chain, noncovalently bound. The interaction between the two chains							
QY	15	is calcium-dependent.							
DB	15	is calcium-dependent.							
QY	15	-1- DOMAIN: Domain B contains 41 X 9 AA tandem repeats. Domains C1							
DB	15	-1- DOMAIN: Domain B contains 41 X 9 AA tandem repeats. Domains C1							
QY	15	and C2 may be involved in membrane binding.							
DB	15	and C2 may be involved in membrane binding.							
QY	15	-1- PTM: Thrombin activates factor V proteolytically to the active							
DB	15	-1- PTM: Thrombin activates factor V proteolytically to the active							
QY	15	cofactor, factor Va (formation of a heavy chain at the N-							
DB	15	cofactor, factor Va (formation of a heavy chain at the N-							
QY	15	terminus and a light chain at the C-terminus).							
DB	15	terminus and a light chain at the C-terminus).							
QY	15	-1- SIMILARITY: Contains 3 PS/8 type A domains.							
DB	15	-1- SIMILARITY: Contains 3 PS/8 type A domains.							
QY	15	-1- SIMILARITY: STRONG, TO COAGULATION FACTOR VIII.							
DB	15	-1- SIMILARITY: STRONG, TO COAGULATION FACTOR VIII.							
QY	15	This SWISS-PROT entry is copyright. It is produced through a collaboration							
DB	15	This SWISS-PROT entry is copyright. It is produced through a collaboration							
QY	15	between the Swiss Institute of Bioinformatics and the EMBL outstation -							
DB	15	between the Swiss Institute of Bioinformatics and the EMBL outstation -							
QY	15	the European Bioinformatics Institute. There are no restrictions on its							
DB	15	the European Bioinformatics Institute. There are no restrictions on its							
QY	15	use by non-profit institutions as long as its content is in no way							
DB	15	use by non-profit institutions as long as its content is in no way							



modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----  
CC EMBL AF191308; AAC28381.1; -  
CC HSSP: P12259; 1CZT.  
DR InterPro: IPR001117; Cu-oxidase.  
DR InterPro: IPR000421; FAS8\_C.  
DR Pfam: PF00754; Cu-oxidase; 3.  
DR SMART: SM00231; FAS8C\_2.  
DR PROSITE: PS01285; FAS8C\_1; 2.  
DR PROSITE: PS01286; FAS8C\_2; 2.  
DR PROSITE: PS00022; FAS8C\_3; 2.  
DR PROSITE: PS00079; MULTICOPPER OXIDASE1; 2.  
KW Blood coagulation; Glycoprotein; Sulfation; Calcium; Signal; Zymogen; Repeat.  
FT SIGNAL 1 22 POTENTIAL.  
FT CHAIN 23 2258 COAGULATION FACTOR V.  
FT CHAIN 23 737 HEAVY CHAIN (BY SIMILARITY).  
FT PEPTIDE 738 1611 ACTIVATION PEPTIDE (CONNECTING REGION)  
(BY SIMILARITY).  
FT CHAIN 1612 2258 LIGHT CHAIN (BY SIMILARITY).  
FT DOMAIN 30 329 F5/8 TYPE A 1.  
FT DOMAIN 203 329 PLASTOCYANIN-LIKE 1.  
FT DOMAIN 348 683 PLASTOCYANIN-LIKE 2.  
FT DOMAIN 535 683 F5/8 TYPE A 2.  
FT DOMAIN 691 1611 PLASTOCYANIN-LIKE 3.  
FT DOMAIN 1168 1539 PLASTOCYANIN-LIKE 4.  
B.  
41 X 9 AA APPROXIMATE TANDEM REPEATS OF  
T-L-S-P-D-L- (GS) (HQ)-T.  
FT REPEAT 1168 1176 1.  
FT REPEAT 1177 1185 2.  
FT REPEAT 1186 1194 3.  
FT REPEAT 1195 1203 4.  
FT REPEAT 1204 1212 5.  
FT REPEAT 1213 1221 6.  
FT REPEAT 1222 1230 7.  
FT REPEAT 1231 1239 8.  
FT REPEAT 1240 1248 9.  
FT REPEAT 1249 1257 10.  
FT REPEAT 1258 1266 11.  
FT REPEAT 1267 1275 12.  
FT REPEAT 1276 1284 13.  
FT REPEAT 1285 1293 14.  
FT REPEAT 1294 1302 15.  
FT REPEAT 1303 1311 16.  
FT REPEAT 1312 1320 17.  
FT REPEAT 1321 1329 18.  
FT REPEAT 1330 1338 19.  
FT REPEAT 1339 1347 20.  
FT REPEAT 1348 1356 21.  
FT REPEAT 1357 1365 22.  
FT REPEAT 1366 1374 23.  
FT REPEAT 1375 1383 24.  
FT REPEAT 1384 1392 25.  
FT REPEAT 1393 1401 26.  
FT REPEAT 1402 1410 27.  
FT REPEAT 1411 1419 28.  
FT REPEAT 1420 1428 29.  
FT REPEAT 1429 1437 30.  
FT REPEAT 1438 1446 31.  
FT REPEAT 1447 1455 32.  
FT REPEAT 1456 1464 33.  
FT REPEAT 1465 1473 34.  
FT REPEAT 1474 1482 35.  
FT REPEAT 1483 1491 36.  
FT REPEAT 1492 1500 37.  
FT REPEAT 1501 1509 38.  
FT REPEAT 1510 1518 39.  
FT REPEAT 1519 1527 40.  
FT REPEAT 1531 1539 41.

FT DOMAIN 1616 1941 F5/8 TYPE A 3.  
FT DOMAIN 1616 1785 PLASTOCYANIN-LIKE 5.  
FT DOMAIN 1795 1941 PLASTOCYANIN-LIKE 6.  
FT DOMAIN 1941 2095 F5/8 TYPE C 1.  
FT DOMAIN 2100 2255 F5/8 TYPE C 2.  
FT SITE 737 738 CLEAVAGE (BY THROMBIN) (BY  
SIMILARITY).  
FT SITE 1029 1030 CLEAVAGE (BY THROMBIN) (BY  
SIMILARITY).  
FT SITE 1611 1612 CLEAVAGE (BY THROMBIN) (BY  
SIMILARITY).  
FT DISULFID 167 193 PROBABLE.  
FT DISULFID 499 525 PROBABLE.  
FT DISULFID 1759 1785 PROBABLE.  
FT DISULFID 1941 2095 BY SIMILARITY.  
FT DISULFID 2100 2255 BY SIMILARITY.  
FT MOD\_RES 692 692 SULFATION (POTENTIAL).  
FT MOD\_RES 696 696 SULFATION (POTENTIAL).  
FT MOD\_RES 724 724 SULFATION (POTENTIAL).  
FT MOD\_RES 726 726 SULFATION (POTENTIAL).  
FT MOD\_RES 745 745 SULFATION (POTENTIAL).  
FT MOD\_RES 1560 1560 SULFATION (POTENTIAL).  
FT MOD\_RES 1576 1576 SULFATION (POTENTIAL).  
FT MOD\_RES 1581 1581 SULFATION (POTENTIAL).  
FT MOD\_RES 1584 1584 SULFATION (POTENTIAL).  
FT MOD\_RES 1588 1588 SULFATION (POTENTIAL).  
FT MOD\_RES 1631 1631 SULFATION (POTENTIAL).  
FT CARBOHYD 153 153 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 225 225 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 239 239 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 297 297 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 382 382 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 460 460 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 467 467 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 553 553 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 741 741 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 752 752 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 760 760 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 776 776 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 782 782 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 899 899 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 960 960 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1048 1048 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1057 1057 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1066 1066 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1073 1073 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1089 1089 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1174 1174 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1480 1480 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1537 1537 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1597 1597 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1737 1737 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1886 1886 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 2044 2044 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 2243 2243 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 2258 AA; 256078 MW; 9159B9E076A2ACC CRC64;

Query Match 5.2%; Score 137; DB 1; Length 2258;  
Best Local Similarity 18.4%; Pred. No. 0.21;  
Matches 96; Conservative 65; Mismatches 171; Indels 190; Gaps 20;

Oy 25 NOADKAOPKSTVDAAKTAANADAAQGEHGLFVIDATVTHAEVPPDR----- 78  
Db 328 NCAKTRPKLTLTDQRHRIKWEFTIAE-----EVLWDYAPILIPANMDKRYRLH 379  
Oy 79 -----DHPARVVVM-----ETVEKVMRLADGVEXQFTFGQGVPMIRVREG 122  
Db 380 LDNFSNQIGKHKKVVKYQYODESFTKRLNPNNKEDS-----ILGPVIRAQVR 428  
Oy 123 DTIEVQFSN-----HPDSKMPHNVDFHAATGPGGAGASTAGHSTBSFKL- 171  
Db 429 DTLKIVFNMASSRSYSIYPHGVTSPYEDDVNSSSTSDNNTMIRAVQGETIYTKWNILE 488

QY 172 --OPGLVYVHCANP-----VGMHANGMYGILVEPEKLPK-----VDKEYVYVMOG 217  
 DB 489 SEPEPNDACCLTRPRYSVNDITRDIASLIGLILICKRSKSLDKGIQRTALIEQKAVYA 548  
 QY 218 DF-YTKGKYEGOGLQPF--DMEKAIREDAYVFNQSVGALTGEMAKKAVETVRLFVG 274  
 DB 549 VEDENKSWYIEIDNIYFCENPEKVKRDKPF-----YESNIMSTINGVYP 593  
 QY 275 NGCPNLTSEFHYIGELFD-----KVHEEGK--GENHNIQTLLIPAG 314  
 DB 594 ESIFPL-----GFCFDDTVQMHGCSVRTHDNILTIHFTGHSFYGKHEDTLTLPFR 646  
 QY 315 GAITEFKVDVPGDYVLVD-----HAIFPAFNKALG 346  
 DB 647 GESVTVMDVNGVTWMLTMSNPRKKQLKFRDYKCIKRDDEDESEIYVBSSTTLT 705  
 QY 347 ILKV-----EGEENHEIYSHK-----QTDVAVL 369  
 DB 706 TRKMHDSSENKEEENDEYDODLASVLGIRSFNSSLVQEDDEFNLTALENSEFI 765  
 QY 370 PEGAPQALITQAPK---TPAPANL-----QEQIKAG 398  
 DB 766 PSTDRADVSNSSSPGNISRAPNFTPEPRKILPHEATKAG 807  
 RESULT 23  
 ID LAC1\_THACU STANDARD; PRT; 576 AA.  
 AC P56193;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE Laccase 1 precursor (EC 1.10.3.2) (Benzenediol: oxygen oxidoreductase)  
 GN (urushiol oxidase) (diphenol oxidase).  
 OS Thanatephorus cucumeris (Black scurf of potato) (Rhizoctonia solani).  
 OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Heterobasidiomycetes;  
 OC Heterobasidiomycetidae; Ceratobasidiales; Ceratobasidiaceae;  
 OC Thanatephorus.  
 OC NCBI\_TaxID=107832;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=R2;  
 RX MEDLINE=96171523; PubMed=8598061;  
 RA Wahlstromer J.A., Xu F., Brown K.M., Brown S.H., Goughly E.J.,  
 RT Haliker T., Kauppinen S., Pederson A., Schneider P.;  
 RT "The identification and characterization of four laccases from the  
 RT plant pathogenic fungus Rhizoctonia solani.";  
 RL Curr. Genet. 29:395-403(1996).  
 CC -1- FUNCTION: LIGNIN DEGRADATION AND DETOXIFICATION OF LIGNIN-DERIVED  
 CC PRODUCTS (PROBABLE). THIS ISOZYME SHOWS OPTIMAL ACTIVITY AT PH 6,  
 CC -1- CATALYTIC ACTIVITY: 4 benzenediol + O(2) = 4 benzoemulquinone + 2  
 CC H(2)O.  
 CC -1- COFACTOR: BINDS 4 CU-IONS PER MOLECULE. THREE DISTINCT CU  
 CC CENTERS KNOWN AS TYPE 1 OR BLUE, TYPE 2 OR NORMAL, AND TYPE  
 CC 3 OR COUPLED BINUCLEAR (BY SIMILARITY).  
 CC -1- SUBUNIT: Homodimer.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- TISSUE SPECIFICITY: IN MYCELA, AT A LOWER LEVEL THAN LC4.  
 CC -1- SIMILARITY: BELONGS TO THE FAMILY OF MULTICOPPER OXIDASES.  
 CC -1- SIMILARITY: Contains 3 plastocyanin-like domains.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation.  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL; 254275; -; NOT\_ANNOTATED\_CDS.  
 CC DR PIR; S68117; S68117.  
 CC DR InterPro; IPR001117; Cu-oxidase.

DR InterPro; IPR002355; MultiCu oxidase2.  
 DR Pfam; PF00394; Cu-oxidase; 3.  
 DR PROSITE; PS00079; MULTICOPPER\_OXIDASE1; 1.  
 DR PROSITE; PS00080; MULTICOPPER\_OXIDASE2; FALSE\_NEG.  
 KW Oxidoreductase; Signal; Copper; Metal-binding; Glycoprotein; Repeat;  
 KW Lignin degradation; Multigene family.  
 FT SIGNAL 19  
 FT CHAIN 20 576  
 FT DOMAIN 21 145  
 FT DOMAIN 157 304  
 FT DOMAIN 7 304  
 FT DOMAIN 82 82  
 FT METAL 84 84  
 FT METAL 127 127  
 FT METAL 129 129  
 FT METAL 471 471  
 FT METAL 474 474  
 FT METAL 476 476  
 FT METAL 523 523  
 FT METAL 524 524  
 FT METAL 525 525  
 FT METAL 529 529  
 FT CARBOHYD 41 41  
 FT CARBOHYD 182 182  
 FT CARBOHYD 228 228  
 FT CARBOHYD 294 294  
 FT CARBOHYD 368 368  
 SQ SEQUENCE 576 AA; 64377 MW; CSEB32918038AB48 CRC64;  
 Query Match 5.1%; Score 133.5; DB 1; Length 576;  
 Best Local Similarity 20.2%; Pred. No. 0.061;  
 Matches 106; Conservative 63; Mismatches 175; Indels 181; Gaps 26;  
 QY 86 VMKEVEKMKRLA-----DGEYQFWTFGGVPGQMRVREGDITVEQFSN---HPDSK 136  
 DB 17 VLAARVEYGLKTSIDELAPDGVKNATLVNGYVGPPLFANQGDILKVKVQKLTNPEWY 76  
 QY 137 MDHNVDFHA-----TGPGGABASFT-----APGHTSFSEFKA-LQPLVYVHCAN 182  
 DB 77 RTTSHMHGLOHRRNADDGP-----SFVTCQPIVRPASTYTTIPLDQGTGYWH--- 127  
 QY 183 APVGMHANGMYG-LIVEPKE--GLPRVDEKEYVM--QGFYTKGKGEGGLQPPDMEX 237  
 DB 128 SHLSQYVDGLGRLVYDPKPHRLYVDDEKTVLLIGDWY-----HESSK 175  
 QY 238 AI-----REDAEYVFNQSVGALTGEN-----ALKKVGETVLFVNGSGPNLT 281  
 DB 176 AIIASGNITRQRPVSATNGK--GRDPDPTPNPDTLYTLKVRGGRVLRVINSSEIAS 234  
 QY 282 SSFHVIGELFDKVFEGKGENHNIQTLLIPAGAAITFEKYDV----- 325  
 DB 235 PRFSV-----EGHKV--TVIADGVSTKRYVDAPDILAGRIDCVLEAN 277  
 QY 326 --PGDYVL-----VDHAIAPAFN---KGA 344  
 DB 278 QEPDVTWNAIPLTVNPNKTAQALVYEDRRRYPHPKGYRKSVEALIKWNHGHKGG 337  
 QY 345 LGILKKEG-----EENHEIYS-----HKQTDVAVLEPGAPQALITQAPTPAPANL 391  
 DB 338 RGLSGHGLKARMIEGSHLSRSVVRKONETTVVMDSESKVPLEYGAAGSKAPADL 397  
 QY 392 QEQIKAGKATVDSNCAHQPDGKGVNAPPL-----ANSDYLNADHARAA 438  
 DB 398 VLDL-----TFGLNFTATGMMI--NGIPYSPKPIPLTKLITDDEGVTSEDFTEEH---T 448  
 QY 439 SIYANGLSKITVNGNGYESVMPALASDOQIANVITYLNSFGN 483  
 DB 449 VILPKKCIKIEFNKSGSIPITHPVHLGH-----TWVVOGNGN 487  
 RESULT 24  
 ID LAC4\_THACU STANDARD; PRT; 531 AA.

AC 002081; 395 395 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE Laccase 4 precursor (EC 1.10.3.2) (Benzenediol:oxygen oxidoreductase)  
 DE (urishiol oxidase) (diphenol oxidase).  
 GN LCC4.  
 OS Thanatephorus cucumeris (Black scurf of potato) (Rhizoctonia solani).  
 OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Heterobasidiomycetes;  
 OC Heterobasidiomycetidae; Ceratobasidiatales; Ceratobasidiaceae;  
 OC Thanatephorus.  
 OC NCBI\_Taxid=107832;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RC STRAIN=RS22;  
 RX MEDLINE=96171523; PubMed=8598061;  
 RA Wahleithner J.A., Xu F., Brown K.M., Brown S.H., Golightly E.J.,  
 RA Halter T., Kauppinen S., Pederson A., Schneider P.,  
 RT "The identification and characterization of four laccases from the  
 RT plant pathogenic fungus Rhizoctonia solani."  
 RL Curr. Genet. 29:395-403 (1996).  
 CC -1- FUNCTION: LIGNIN DEGRADATION AND DETOXIFICATION OF LIGNIN-DERIVED  
 CC PRODUCTS (PROBABLE). THIS ISOZYME SHOWS OPTIMAL ACTIVITY AT PH 7.  
 CC -1- CATALYTIC ACTIVITY: 4 benzenediol + O(2) = 4 benzoemquinone + 2  
 CC H(2)O.  
 CC -1- COFACTOR: BINDS 4 CU-IONS PER MOLECULE. THREE DISTINCT CU  
 CC CENTERS KNOWN AS TYPE 1 OR BLUE, TYPE 2 OR NORMAL, AND TYPE  
 CC 3 OR COUPLED BINUCLEAR (BY SIMILARITY).  
 CC -1- SUBUNIT: Homodimer.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- TISSUE SPECIFICITY: IN MYCELIA, AT A HIGHER LEVEL THAN LCC1, LCC2  
 CC AND LCC3.  
 CC -1- SIMILARITY: BELONGS TO THE FAMILY OF MULTICOPPER OXIDASES.  
 CC -1- SIMILARITY: Contains 3 plastocyanin-like domains.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; 254277; CAA91042.1; -  
 CC PIR; S68120; S68120.  
 DR InterPro: IPR001117; Cu-oxidase.  
 DR InterPro: IPR002355; MulticCu\_oxidase2.  
 DR Pfam; PF00394; Cu-oxidase; 3.  
 DR PROSITE; PS00079; MULTICOPPER OXIDASE1; FALSE NEG.  
 DR PROSITE; PS00080; MULTICOPPER OXIDASE2; FALSE NEG.  
 KM Oxidoreductase; Signal; Copper; Metal-binding; Glycoprotein; Repeat;  
 KM Lignin degradation; Multigene family; Polymorphism.  
 FT SIGNAL 1 19  
 FT CHAIN 20 531  
 FT DOMAIN 23 146  
 FT DOMAIN 158 315  
 FT DOMAIN 384 507  
 FT METAL 83 83  
 FT METAL 85 85  
 FT METAL 128 128  
 FT METAL 130 130  
 FT METAL 427 427  
 FT METAL 430 430  
 FT METAL 432 432  
 FT METAL 439 439  
 FT METAL 479 479  
 FT METAL 480 480  
 FT METAL 481 481  
 FT METAL 484 484  
 FT CARBOHYD 66 66  
 FT CARBOHYD 109 109  
 FT CARBOHYD 186 186  
 FT CARBOHYD 231 231  
 FT CARBOHYD 280 280

FT CARBOHYD 395 395 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT VARIANT 42 42 P -> S.  
 FT VARIANT 119 119 H -> R.  
 FT VARIANT 246 246 R -> S.  
 FT VARIANT 256 256 P -> L.  
 FT VARIANT 261 261 P -> A.  
 SQ SEQUENCE 531 AA; 57545 MW; 737339803F75AB19 CRC64;  
 Query Match  
 Best Local Similarity 20.3%; Score 131; DB 1; Length 531;  
 Matches 93; Conservative 54; Mismatches 20; Indels 108; Gaps 17;  
 QY 59 LPVIDAIVTHAEVPPVDRHPAKVVK--METVEKYMRADGVYQFWTGGVPGQM 116  
 DB 8 LPLAAVST-----PAPAAYRNYKFDIKNNVADPGRPISVNGVLPGTL 54  
 QY 117 IRRRGDITIEVQFSNH---PDSKMHNVDFH---AATGPGGAASFTAP-GHSTSEF 168  
 DB 55 ITANKGDTLRINVTQLTDPSPKRRATTHMGFLFOATTADBDGPAFVTCPIAQNLSYTY 114  
 QY 169 KAL---OPGLVYHCAVAPVGMHNGMYG-LIVER---KEGPKYDKEXYVWQDPY 220  
 DB 115 ELPFLHGOTGMVYH---AHLSQYVDGIRGLVITDPPDKSRVDVDASTVWLEBDY 171  
 QY 221 -----TKKXGEGG---LQPFMEKAIREDAEYVFNVSVALTGENALAKVGETVR 270  
 DB 172 HTPAPVLEKQMFSTNNLTALLSPDPSGLINGKRVY-----CGPAVPSVIVWKKGRYR 226  
 QY 271 LFGVNGGNLTSSFFVIGEIPDKVHFBEGKGNNHIOITLLIPAGAAITFEFVDVPGDY 330  
 DB 227 LRVINASAIGSTFPIBGRRLVLEADG-----IPHQLPVDSPFOIVAGORYS 274  
 QY 331 LVDAHIFAPFKNGALGIKVEGENHEIYSHKQTDAYVLPGAPQAIPTQE----- 381  
 DB 275 VIVEANQTAANYIRAPVTWAGATNANLDPNNVFAVLHBCAPAEPTTEGSAIGTAL 334  
 QY 382 -----APKTPAPANLOEQIRAKATYDS-----NCAACHOPDQKGVNAPPP 423  
 DB 335 VEENHLALINGAPGSGAPADVSINLAIGRSVGDILFTFNINIXEAP-----SLPT 387  
 QY 424 LANSDYLNADHARAASIVANGLSKITVNGQYESVMP 461  
 DB 388 L-----LKIANNASNDADPTPNEHTVLP 412  
 RESULT 25  
 FAS\_HUMAN STANDARD; PRT; 2224 AA.  
 AC P12259; Q14285;  
 DT 01-OCT-1989 (Rel. 12, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Coagulation factor V precursor (Activated protein C cofactor).  
 GN F5.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OC NCBI\_Taxid=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND VARIANT MET-1764.  
 RX MEDLINE=92232668; PubMed=1567832;  
 RA Cripe L.D., Moore K.D., Kane W.H.;  
 RT "Structure of the gene for human coagulation factor V";  
 RL Biochemistry 31:3777-3785 (1992).  
 RN [2]  
 RP SEQUENCE FROM N.A., AND VARIANT GLU-925.  
 RX MEDLINE=87260886; PubMed=3110773;  
 RA Jenny R.J., Pittman D.D., Toole J.J., Kitz R.W., Aldape R.A.,  
 RA Hewick R.M., Kaufman R.J., Mann K.G.;  
 RT "Complete cDNA and derived amino acid sequence of human factor V";  
 RL Proc. Natl. Acad. Sci. U.S.A. 84:4846-4850 (1987).  
 RN [3]  
 RP SEQUENCE OF 1-1600 FROM N.A., AND VARIANTS GLU-925 AND ILE-1285.

RA MEDLINE=8107560; PubMed=2827731;  
RA Kane W.H., Ichinose A., Hagen F.S., Davie E.W.;  
RT "Cloning of cDNAs coding for the heavy chain region and connecting  
RT region of human factor V, a blood coagulation factor with four types  
RT of internal repeats.";   
RA Biochemistry 26:6508-6514(1987).  
RN [4]  
RP SEQUENCE OF 1188-1215 AND 1315-2224 FROM N.A.  
RX MEDLINE=86313665; PubMed=3092220;  
RA Kane W.H., Davie E.W.;  
RT "Cloning of a cDNA coding for human factor V, a blood coagulation  
RT factor homologous to factor VIII and ceruloplasmin.";   
RA Proc. Natl. Acad. Sci. U.S.A. 83:6800-6804(1986).  
RN [5]  
RP PARTIAL SEQUENCE FROM N.A.  
RC TISSUE=Fibroblast;  
RX MEDLINE=9303619; PubMed=8454869;  
RA Shen N.L., Fan S.-T., Pyati J., Graff R., Lapolla R.J.,  
RA Edgington T.S.;  
RT "The serine protease cofactor factor V is synthesized by  
RT lymphocytes.";   
RA J. Immunol. 150:2992-3001(1993).  
RN [6]  
RP SULFATION.  
RX MEDLINE=94264012; PubMed=8204629;  
RA Pittman D.D., Tomkinson K.N., Michnick D., Seligsohn U.,  
RA Kaufman R.J.;  
RT "Posttranslational sulfation of factor V is required for efficient  
RT thrombin cleavage and activation and for full procoagulant activity.";   
RA Biochemistry 33:6952-6959(1994).  
RN [7]  
RP SULFATION.  
RX MEDLINE=90366699; PubMed=2168225;  
RA Horton G.L.;  
RT "Sulfation of tyrosine residues in coagulation factor V.";   
RA Blood 76:946-952(1990).  
RN [8]  
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 2065-2224.  
RX MEDLINE=20052169; PubMed=1058686;  
RA Macedo-Ribeiro S., Bode W., Huber R., Quinn-Alten M.A., Kim S.W.,  
RA Ortel T.L., Bourenkov G.P., Bartunik H.D., Stubbs M.T., Kane W.H.,  
RA Fuentes-Prior P.;  
RT "Crystal structures of the membrane-binding C2 domain of human  
RT coagulation factor V.";   
RA Nature 402:434-439(1999).  
RN [9]  
RP VARIANT MET-1764.  
RX MEDLINE=95179146; PubMed=7874144;  
RA Bayeson T.A., Ireland H., Olds R.J., Thein S.L., Lane D.A.;  
RT "A polymorphism in the human coagulation factor V gene.";   
RA Hum. Mol. Genet. 3:2085-2085(1994).  
RN [10]  
RP VARIANT APCR GLN-534.  
RX MEDLINE=94217810; PubMed=8164741;  
RA Bertina R.M., Koelmaan B.P.C., Koster T., Rosendaal F.R.,  
RA Dirven R.J., de Ronde H., van der Velden P.A., Rietma P.H.;  
RT "Mutation in blood coagulation factor V associated with resistance to  
RT activated protein C.";   
RA Nature 369:64-67(1994).  
RN [11]  
RP VARIANTS ILE-1285 AND ARG-1327.  
RX MEDLINE=96351768; PubMed=8713778;  
RA Lunghi B., Iacovello L., Gemmati D., Dilasio M.G., Castoldi E.,  
RA Pinotti M., Castellani G., Redaelli R., Mariani G., Marchetti G.,  
RA Bernardi F.;  
RT "Detection of new polymorphic markers in the factor V gene:  
RT association with factor V levels in plasma.";   
RA Thromb. Haemost. 75:45-48(1996).  
RN [12]  
RP VARIANT APCR GLY-334, AND VARIANT LYS-513.  
RX MEDLINE=98122763; PubMed=9454741;  
RA Chan W.P., Lee C.K., Kwong Y.L., Lam C.K., Liang R.;  
RT "A novel mutation of Arg306 of factor V gene in Hong Kong Chinese.";

RL Blood 91:1135-1139(1998).  
RN [13]  
RP VARIANT APCR THR-334.  
RX MEDLINE=98122764; PubMed=9454742;  
RA Williamson D., Brown K., Luddington R., Baglin C., Baglin T.;  
RT "Factor V Cambridge: a new mutation (Arg306-to-Thr) associated with  
RT resistance to activated protein C.";   
RA Blood 91:1140-1144(1998).  
RN [14]  
RP VARIANTS HIS-107; THR-413; LYS-513; SER-809; THR-817; ARG-858;  
RX ARG-865; GLU-925; GLN-1146; ALA-1530; SER-1685; VAL-1749; MET-1764;  
RP ILE-1820 AND GLY-2222, AND VARIANT APCR GLN-534.  
RX MEDLINE=99318093; PubMed=10391209;  
RA Cargill M., Altschuler D., Ireland J., Sklar P., Ardlie K., Patil N.,  
RA Shaw N., Lane C.R., Lim E.P., Kalyanaraman N., Nemesh J., Ziaugra L.,  
RA Friedland L., Rolfe A., Warrington J., Lipshutz R., Daley G.O.,  
RA Lander E.S.;  
RT "Characterization of single-nucleotide polymorphisms in coding regions  
RT of human genes.";   
RA Nat. Genet. 22:231-238(1999).  
RN [15]  
RP ERRATUM.  
RA Cargill M., Altschuler D., Ireland J., Sklar P., Ardlie K., Patil N.,  
RA Shaw N., Lane C.R., Lim E.P., Kalyanaraman N., Nemesh J., Ziaugra L.,  
RA Friedland L., Rolfe A., Warrington J., Lipshutz R., Daley G.O.,  
RA Lander E.S.;  
RT "Characterization of single-nucleotide polymorphisms in coding regions  
RT of human genes.";   
RA Nat. Genet. 22:231-238(1999).  
RN [15]  
RP -1- FUNCTION: COAGULATION FACTOR V IS A COFACTOR THAT PARTICIPATES  
CC WITH FACTOR XA TO ACTIVATE PROTHROMBIN TO THROMBIN.  
CC -1- SUBUNIT: Factor Va is composed of a heavy chain and a light  
CC chain, noncovalently bound. The interaction between the two chains  
CC is calcium-dependent.  
CC -1- DOMAIN: DOMAIN B CONTAINS 35 X 9 AA TANDEM REPEATS, AND 2 X 17 AA  
CC REPEATS.  
CC -1- PTM: Thrombin activates factor V proteolytically to the active  
CC cofactor, factor Va (formation of a heavy chain at the N-  
CC terminus and a light chain at the C-terminus).  
CC -1- PTM: SULFATION IS REQUIRED FOR EFFICIENT THROMBIN CLEAVAGE AND  
CC ACTIVATION AND FOR FULL PROCOAGULANT ACTIVITY.  
CC -1- DISEASE: Defects in F5 are the cause of Owren parahemophilia  
CC (MM:227400), an hemorrhagic diathesis.  
CC -1- DISEASE: Defects in F5 are the cause of resistance to activated  
CC protein C (APCR) [MM:188055], a form of thrombophilia. The APCR  
CC mutation is found in about 5% of the population which suggest that  
CC a slight thrombotic tendency may confer some advantage in fetal  
CC implantation.  
CC -1- SIMILARITY: Contains 3 F5/8 type A domains.  
CC -1- SIMILARITY: Contains 2 F5/8 type C domains.  
CC -1- SIMILARITY: STRONG, TO COAGULATION FACTOR VIII.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
DR EMBL; L32779; AAB59401.1; -;  
DR EMBL; L32755; AAB59401.1; JOINED.  
DR EMBL; L32756; AAB59401.1; JOINED.  
DR EMBL; L32757; AAB59401.1; JOINED.  
DR EMBL; L32758; AAB59401.1; JOINED.  
DR EMBL; L32759; AAB59401.1; JOINED.  
DR EMBL; L32760; AAB59401.1; JOINED.  
DR EMBL; L32761; AAB59401.1; JOINED.  
DR EMBL; L32762; AAB59401.1; JOINED.  
DR EMBL; L32763; AAB59401.1; JOINED.  
DR EMBL; L32764; AAB59401.1; JOINED.  
DR EMBL; L32765; AAB59401.1; JOINED.  
DR EMBL; L32766; AAB59401.1; JOINED.  
DR EMBL; L32767; AAB59401.1; JOINED.  
DR EMBL; L32768; AAB59401.1; JOINED.



```

CC METAL 142 142 COPPER (TYPE 3) (BY SIMILARITY)
CC METAL 144 144 COPPER (TYPE 3) (BY SIMILARITY)
CC METAL 458 458 COPPER (TYPE 2) (BY SIMILARITY)
CC METAL 461 461 COPPER (TYPE 2) (BY SIMILARITY)
CC METAL 463 463 COPPER (TYPE 3) (BY SIMILARITY)
CC METAL 514 514 COPPER (TYPE 3) (BY SIMILARITY)
CC METAL 515 515 COPPER (TYPE 3) (BY SIMILARITY)
CC METAL 516 516 COPPER (TYPE 3) (BY SIMILARITY)
CC METAL 520 520 COPPER (TYPE 1) (BY SIMILARITY)
CC METAL 525 525 COPPER (TYPE 1) (BY SIMILARITY)
SQ SEQUENCE 533 AA; 58328 MW; ED1570C9E9B1C135 CRC64;

Query Match 4.7%; Score 124; DB 1; Length 533;
Best Local Similarity 22.1%; Pred. No. 0.26;
Matches 64; Conservative 51; Mismatches 131; Indels 44; Gaps 14;

QY 72 VPPVDRDHPAKVAVVMEVKEVYKRLADGVEYQFWTFGGQVPGQMTIRREGDTIEVQFSN 131
DB 36 IEPPLQDPANGLNINLQIGSVV--LPSATQTWGNGNLGPALRLRGKAVTIDITN 93
QY 132 H-PDSMPHNVDFHAATGCG--GGAASFTAGHTSTSFYKALQCG--LYYHCVAAY 185
DB 94 ALPEATVH--WHGLEIPEGVGQQA-LIQGAKROVTFAVEQPAATCWFPHTHSKT 149
QY 186 GMIHANGMYGLIIVEPKG---LPRK---VDKEYVMQDFFYKYGEGLOPFMEKA 238
DB 150 GHQVANGGLGLVILIDSSETLPLPKMGVDDIPLVLDKLT--DHGOVDIQLDVTAA 207
QY 239 IREDAEYVFNQ-----SVGALTGENA---LKAVGETVRLFVNGNG--- 277
DB 208 VGMFGGRMLTNGVYPOQTIRGVMVRLRLNGNANSLNALSDBGPMYVADGGLAE 267
QY 278 PNTSSFFHYI-GEIPDKVHEGCKENHNIQTLLPAGCAATIEFKVDP 326
DB 268 PUVVRELPLIMGERFE-VIVDRDQSLDLVTLPLVQMGTLAPFDQPLP 316

RESULT 27
LAC3_THACU STANDARD; PRT; 572 AA.
AC 002079;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Laccase 3 precursor (EC 1.10.3.2) (Benzene:diol:oxygen oxidoreductase)
DE (Urishiol oxidase) (Diphenol oxidase).
GN LCC3.
OS Thanaeophorus cucumeris (Black scurf of potato) (Rhizoctonia solani).
OS Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Heterobasidiomycetes;
OC Heterobasidiomycetidae; Ceratobasidiales; Ceratobasidiaceae;
OC Thanaeophorus.
OX NCBI_TaxID=107832;
RX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RS22;
RX MEDLINE=96171523; PubMed=8598061.
RA Wahlström J.A., Xu F., Brown K.M., Brown S.H., Golightly E.J.,
RA Halckier T., Kampunen S., Pederson A., Schneider P.;
RA "The identification and characterization of four laccases from the
RA plant pathogenic fungus Rhizoctonia solani.";
RL Curr. Genet. 29:395-403 (1996).
CC -1- FUNCTION: LIGNIN DEGRADATION AND DETOXIFICATION OF LIGNIN-DERIVED
CC PRODUCTS (PROBABLE).
CC -1- CATALYTIC ACTIVITY: 4 benzene:diol + O(2) = 4 benzosemiquinone + 2
CC H(2)O.
CC -1- COFACTOR: BINDS 4 CU-IONS PER MOLECULE. THREE DISTINCT CU
CC CENTERS KNOWN AS TYPE 1 OR BLUE, TYPE 2 OR NORMAL, AND TYPE
CC 3 OR COUPLED BINUCLEAR (BY SIMILARITY).
CC -1- SUBUNIT: Homodimer (By similarity).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: IN MYCELA, AT A LOWER LEVEL THAN LCC4.
CC -1- SIMILARITY: BELONGS TO THE FAMILY OF MULTICOPPER OXIDASES.
CC -1- SIMILARITY: Contains 3 plastocyanin-like domains.

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: 254215; CAA90942.1; -
CC PIR: S68119; S68119.
CC InterPro: IPR001117; Cu-oxidase.
CC InterPro: IPR002355; MultiCu_oxidase2.
CC Pfam: PF00394; Cu-oxidase; 3.
CC PROSITE: PS00079; MULTICOPPER_OXIDASE1; 1.
CC PROSITE: PS00080; MULTICOPPER_OXIDASE2; FALSE_NEG.
CC Oxidoreductase; Signal; Copper; Metal-binding; Glycoprotein; Repeat;
CC Lignin degradation; Multigene family; Polymorphism.
CC SIGNAL 1 18
CC CHAIN 19 572
CC DOMAIN 21 145
CC DOMAIN 422 540
CC METAL 82 82
CC METAL 84 84
CC METAL 127 127
CC METAL 129 129
CC METAL 470 470
CC METAL 473 473
CC METAL 475 475
CC METAL 522 522
CC METAL 523 523
CC METAL 524 524
CC METAL 526 526
CC CARBOHYD 162 162
CC CARBOHYD 228 228
CC CARBOHYD 294 294
CC CARBOHYD 367 367
CC CARBOHYD 405 405
CC VARIANT 159 159
CC VARIANT 359 359
CC VARIANT 418 418
CC VARIANT 448 448
SQ SEQUENCE 572 AA; 63747 MW; DFL1E2AF9F108CE4 CRC64;

Query Match 4.7%; Score 124; DB 1; Length 572;
Best Local Similarity 21.4%; Pred. No. 0.28;
Matches 79; Conservative 55; Mismatches 124; Indels 112; Gaps 20;

QY 86 VMEVTEVKEVWRLA-----DGEYQFWTFGGQVPGQMTIRREGDTIEVQFSN---HPDSK 136
DB 17 VLARIVEYVTKISNGKIPADGVERDALTVMGYPGLIFANKGDLTKVAVQKLNPDWY 76
QY 137 MBHNVDFHAA-----TGGGGAASFTAGHTSTSEFK-ALPGLVYHCVAAYVCM 187
DB 77 RTTSHHGLQHRNADDGPAFTVQCPI-VQASATYTMPLDGTGTWYH---SHLS 132
QY 188 HANGMYG-LIIVEPKGPKV---DKEYVMQDFFY-TKKR---YEOGLQPPDMK 237
DB 133 QYVDGIRGLVLYDPDPHRLVYDIDDEKTVLIDGWHTSSKAILATNTITLQOPDS-- 190
QY 238 AIREDAEYVFNQSGVLTGEN-----ALKAVGETVRLFVNGGPNLTSSFFHYIG 288
DB 191 -----ATNGK-GRPPDVTTPANPNTLYLTKYKRGKRYLRYVNS--SALASF----- 235
QY 289 EIFDKVHEGCKENHNIQTLLIPAGGAATIEFKVDVPGDYVLVDHAFRAFNKAGLIL 348
DB 236 ---RMSIGHK-----MTVIADGVSTKRYQVD----- 260
QY 349 KYEGENHEIYSHKQTDVAVLEBGAFOALDIOEAPCT---PAP-ANLQEOIKAGKATYDS 404
DB 261 -----SPDLAGORIDAV-----VEANQEPDTWINAPLTVNANKTAQALLIYED 305

```

QY 405 NCACHQPDG 414  
 Db 306 DRRPYHPKG 315

RESULT 28  
 FETS YEAST  
 ID FETS YEAST STANDARD; PRT; 622 AA.  
 AC P43561;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Iron transport multicopper oxidase FETS precursor (EC 1.10.3.2).  
 GN FETS OR YFL041W.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_TaxId=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=5288C / AB972;  
 RX MEDLINE=95400292; PubMed=7670463;  
 RA Murakami Y., Naitou M., Hagiwara H., Shibata T., Ozawa M.,  
 RA Yamazaki M., Tashiro H., Eki T.;  
 RT "Analysis of the nucleotide sequence of chromosome VI from  
 RT Saccharomyces cerevisiae.";  
 RL Nat. Genet. 10:261-268(1995).  
 RN [2]  
 RP CHARACTERIZATION.  
 RX MEDLINE=98075360; PubMed=9413439;  
 RA Spitzo T., Byersdorfer C., Duesterhoeft S., Eide D.;  
 RT "The yeast FETS gene encodes a FET3-related multicopper oxidase  
 RT implicated in iron transport.";  
 RL Mol. Genet. 256:547-556(1997).  
 CC -1- FUNCTION: IRON TRANSPORT MULTICOPPER OXIDASE, WHICH IS REQUIRED  
 CC FOR FERROUS IRON HIGH AFFINITY UPTAKE. MAY BE REQUIRED TO OXIDIZE  
 CC FE(II) AND RELEASE IT FROM THE TRANSPORTER. ESSENTIAL COMPONENT OF  
 CC COPPER-DEPENDENT IRON TRANSPORT.  
 CC -1- COPACITOR: THIS PROTEIN BELONGS TO THE MULTICOPPER OXIDASES WHICH  
 CC CONTAIN THREE DISTINCT CU CENTERS KNOWN AS TYPE 1 OR BLUE, TYPE  
 CC 2 OR NORMAL, AND TYPE 3 OR COUPLED BINUCLEAR (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE MULTICOPPER OXIDASE FAMILY.  
 CC -1- SIMILARITY: CONTAINS 3 PLASTOCYANIN-LIKE DOMAINS.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL, D50617, BAA09199.1; -  
 DR PIR, S56214; S56214.  
 DR HSSP, P37064; IASO.  
 DR SGD, S0001853; FETS.  
 DR GO, GO:0004323; F:multicopper ferroxidase iron transport medi. . .; IGI.  
 DR GO, GO:0006826; P:iron ion transport; IMP.  
 DR InterPro, IPR001117; Cu-oxidase.  
 DR InterPro, IPR002355; MultiCu\_oxidase2.  
 DR Pfam, PF00394; Cu-oxidase; 3.  
 DR PROSITE, PS00079; MULTICOPPER OXIDASE1; 1.  
 DR PROSITE, PS00080; MULTICOPPER OXIDASE2; 1.  
 KW Glycoprotein; Transmembrane; Signal; Transport; Oxidoreductase;  
 KW Copper; Repeat; Metal-binding; Iron transport.  
 FT SIGNAL 1  
 FT CHAIN 18  
 FT DOMAIN 19 622 IRON TRANSPORT MULTICOPPER OXIDASE FETS.  
 FT TRANSMEM 574 594 EXTRACELLULAR (POTENTIAL).  
 FT DOMAIN 595 622 POTENTIAL.  
 FT DOMAIN 43 146 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 192 301 PLASTOCYANIN-LIKE 1.  
 FT PLASTOCYANIN-LIKE 2.

FT DOMAIN 392 514 PLASTOCYANIN-LIKE 3.  
 FT METAL 79 79 COPPER (TYPE 2) (BY SIMILARITY).  
 FT METAL 81 81 COPPER (TYPE 2) (BY SIMILARITY).  
 FT METAL 128 128 COPPER (TYPE 3) (BY SIMILARITY).  
 FT METAL 130 130 COPPER (TYPE 3) (BY SIMILARITY).  
 FT METAL 418 418 COPPER (TYPE 1) (BY SIMILARITY).  
 FT METAL 421 421 COPPER (TYPE 2) (BY SIMILARITY).  
 FT METAL 423 423 COPPER (TYPE 3) (BY SIMILARITY).  
 FT METAL 496 496 COPPER (TYPE 1) (BY SIMILARITY).  
 FT METAL 497 497 COPPER (TYPE 1) (BY SIMILARITY).  
 FT METAL 498 498 COPPER (TYPE 3) (BY SIMILARITY).  
 FT METAL 502 502 COPPER (TYPE 1) (BY SIMILARITY).  
 FT CARBOHYD 24 24 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 86 86 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 115 115 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 200 200 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 246 246 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 295 295 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 364 364 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 455 455 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 622 AA; 70880 MW; 961757A6C86B7EAF CRC64;

Query Match 4.5%; Score 118; DB 1; Length 622;  
 Best Local Similarity 20.9%; Pred. No. 0.83;  
 Matches 93; Conservative 54; Mismatches 171; Indels 126; Gaps 22;

QY 99 DGV-EYQWTFGGQVPGMIRVREGDTIEVQFSNHPDSKMPHNVDFA----- 145  
 Db 35 DGLHEKRNIGNGEWPDPDINHEKGRVLELYTNGFQDNTATSLFHLFQNTSLGANOLO 94  
 QY 146 ATGPAGAGASFTAPGHTSTESFKA-LQPGYVYCAVAPGMIANGMYGLIIVEPKG 204  
 Db 95 MDGPEMTVQCPL-VPGQYLYNFTYPRQVGFVFWH--AHNGAQVGDGMRGAFIHDPE 150  
 QY 205 LPRVDKEYVYVNGDFY-----TK--GRYGEQGLQPFDMKALREDAEVVNGSVGA 254  
 Db 151 PFEYDHERVITLSHYHENYKVTVEFLSRYPNTGAEPI-----PQNLFNNTVMV 201  
 QY 255 LTGNAALAKXGET-----VRLFVGN-----GGPNLTSSHVIGE 289  
 Db 202 ----TLDFTEYLFREFLVNGLFVSOYILLEHMSIVEVDGYVKNPFI----- 248  
 QY 290 IFDKVHFEKGKGNHNITTLIPAGAAITFEKVDPGVYLVDAIFRAEFGKALGLK 349  
 Db 249 --DSYLSAGR-----MSVLKAKDKMPTR-----NYAMQIMDETLDVVP 289  
 QY 350 VEGEENHEI---YHKQTDVAVLLPEGAPQALDTGEAPKTPAPANLQEQIKAG-KATYDSN 405  
 Db 290 PELQINQTIQMRYGHS-----LPEARALNIEDCDLDRATNDFLEPLIERDLAHYD-- 341  
 QY 406 CAACHQ-----PDGKGVNAPFPPLANSYLVNADHRAASIVANGSGKITVNGNOYE 457  
 Db 342 ---HQIVMDVRMVLGGVRYAF--FNNITVYVPKVEPLTLTLT---SGKLASDPRIYG 392  
 QY 458 SVMPAIALSDQIANVITYTNSF 481  
 Db 393 DNINAQLKNDIIEV---LNNY 413

RESULT 29  
 LAC1 AGABI STANDARD; PRT; 520 AA.  
 ID LAC1 AGABI  
 AC O12541;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE Laccase I precursor (EC 1.10.3.2) (Benzenediol: oxygen oxidoreductase)  
 GN (Utrishol oxidase) (Diphenol oxidase).  
 DE LCCI  
 OS Agaricus bisporus (Common mushroom).  
 OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;  
 OC Agaricales; Agaricaceae; Agaricus.

OX NCBI\_TaxID=5341;  
 RN SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RC STRAIN=D649; TISSUE=Mycelium;  
 RA MEDLINE=93367392; PubMed=8360614;  
 RT Perry C.R., Smith M., Britnell C.H., Wood D.A., Thurston C.F.;  
 RT "Identification of two laccase genes in the cultivated mushroom  
 Agaricus bisporus";  
 RL J. Gen. Microbiol. 139:1209-1218(1993).  
 CC -1- FUNCTION: LIGNIN DEGRADATION AND DETOXIFICATION OF LIGNIN-DERIVED  
 CC PRODUCTS (PROBABLE).  
 CC -1- CATALYTIC ACTIVITY: 4 benzenediol + O(2) = 4 benzoquinone + 2  
 CC H(2)O.  
 CC -1- COFACTOR: BINDS 4 CU-IONS PER MOLECULE. THREE DISTINCT CU  
 CC CENTERS KNOWN AS TYPE 1 OR BLUE, TYPE 2 OR NORMAL, AND TYPE  
 CC 3 OR COUPLED BINUCLEAR (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- SIMILARITY: BELONGS TO THE FAMILY OF MULTICOPPER OXIDASES.  
 CC -1- SIMILARITY: Contains 3 plastocyanin-like domains.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; L10664; AAC18877.1; -.  
 DR HSSP; P37064; 1A0Z.  
 DR InterPro; IPR001117; Cu-oxidase.  
 DR InterPro; IPR002355; MultiCu\_oxidase2.  
 DR Pfam; PF00394; Cu-oxidase; 3.  
 DR PROSITE; PS00079; MULTICOPPER OXIDASE1; 2.  
 DR PROSITE; PS00080; MULTICOPPER OXIDASE2; 1.  
 KM Oxidoreductase; Signal; Copper; Metal-binding; Lignin degradation;  
 KM Glycoprotein; Repeat; Multigene family.  
 FT SIGNAL 1 19  
 FT CHAIN 20 520  
 FT DOMAIN 21 145 PLASTOCYANIN-LIKE 1.  
 FT DOMAIN 157 305 PLASTOCYANIN-LIKE 2.  
 FT DOMAIN 375 488 PLASTOCYANIN-LIKE 3.  
 FT METAL 82 82 COPPER (TYPE 2) (BY SIMILARITY).  
 FT METAL 84 84 COPPER (TYPE 2) (BY SIMILARITY).  
 FT METAL 127 127 COPPER (TYPE 3) (BY SIMILARITY).  
 FT METAL 129 129 COPPER (TYPE 3) (BY SIMILARITY).  
 FT METAL 417 417 COPPER (TYPE 1) (BY SIMILARITY).  
 FT METAL 420 420 COPPER (TYPE 2) (BY SIMILARITY).  
 FT METAL 422 422 COPPER (TYPE 3) (BY SIMILARITY).  
 FT METAL 470 470 COPPER (TYPE 3) (BY SIMILARITY).  
 FT METAL 471 471 COPPER (TYPE 1) (BY SIMILARITY).  
 FT METAL 472 472 COPPER (TYPE 3) (BY SIMILARITY).  
 FT METAL 476 476 COPPER (TYPE 1) (BY SIMILARITY).  
 FT CARBOHYD 108 108 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 239 239 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 299 299 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 454 454 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 492 492 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 520 AA; 58099 MW; 943DB3F23297B891 CRC64;  
 Query Match 4.4%; Score 116; DB 1; Length 520;  
 Best Local Similarity 26.6%; Pred. No. 0.91; Indels 42; Gaps 13;  
 Matches 59; Conservative 28; Mismatches 93;  
 82 AKVVVKKEETVEKWN---RLA-DGVEYQFWTFGGVPGQMIRREGDTIEVQFSNHPDS-K 136  
 Db 15 SSVAKRTPTFDLVNRLAPDGFERDTVVINGEPGLIIVNKGDSVRIPLNKLISPT 74  
 Qy 137 MPNVDH-----AATGGGGAASFT-----APGHTSTSEK-ALDPGLVYHCAVAP 184  
 Db 75 MRSVSIHMHGFPQARTSGQDG--PSFVNQCPQPPNPTFTVEFSVAESGTFWYH---SH 129  
 Qy 185 VGMHIANGMVGLILV---EPKEGLPKVDK-----YYWQGDPTTNGKKGEGQLQ 231

Db 130 LSTQYCDGLRGAFIYVYDPRDLRHLXYDVDESTVITLAEMHILAPD--ATNEFFSGII 187  
 Qy 222 PFDMAKRAEDAEYVNGSVGALGEMALAKVETVRLV 273  
 Db 188 PQVDGSLINGKGR---FNG--GPLTPFAVVNVVRGKRYRLV 224  
 RESULT 30  
 NOSZ\_PARDE STANDARD; PRT; 652 AA.  
 ID NOSZ\_PARDE  
 AC Q51705;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Nitrous-oxide reductase precursor (EC 1.7.99.6) (N(2)OR).  
 GN NOSZ.  
 OS Paracoccus denitrificans.  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacteriales;  
 OC Rhodobacteraceae; Paracoccus.  
 OX NCBI\_TaxID=266;  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=NL188944;  
 RX MEDLINE=94062841; PubMed=8243476;  
 RA Hoeren F.U., Berke B.C., Ferguson S.J., McCarthy J.E.G.;  
 RT "Sequence and expression of the gene encoding the respiratory  
 RT nitrous-oxide reductase from Paracoccus denitrificans. New and  
 RT conserved structural and regulatory motifs".  
 RL Eur. J. Biochem. 218:49-57(1993).  
 CC -1- FUNCTION: NITROUS-OXIDE REDUCTASE IS PART OF A BACTERIAL  
 CC RESPIRATORY SYSTEM WHICH IS ACTIVATED UNDER ANAEROBIC  
 CC CONDITIONS IN THE PRESENCE OF NITRATE OR NITROUS OXIDE.  
 CC -1- CATALYTIC ACTIVITY: N(2) + H(2)O + acceptor = nitrous oxide +  
 CC reduced acceptor.  
 CC -1- COFACTOR: COPPER (BY SIMILARITY).  
 CC -1- SUBUNIT: HOMODIMER; EACH SUBUNIT CONTAINS 2 DINUCLEAR COPPER  
 CC CENTERS A AND Z. Z IS THOUGHT TO BE THE SITE OF NITROUS OXIDE  
 CC REDUCTION (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: Periplasmic.  
 CC -1- SIMILARITY: IN THE C-TERMINAL SECTION, TO MITOCHONDRIAL AND  
 CC BACTERIAL COX2 SUBUNTS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; X74792; CAA52798.1; -.  
 DR PIR; S39409; S39409.  
 KM Signal; Oxidoreductase; Copper; Metal-binding; Periplasmic.  
 FT SIGNAL 1 57  
 FT CHAIN 58 652  
 FT METAL 595 595 NITROUS-OXIDE REDUCTASE.  
 FT METAL 630 630 COPPER A1 (BY SIMILARITY).  
 FT METAL 630 630 COPPER A1 (BY SIMILARITY).  
 FT METAL 630 630 COPPER A2 (BY SIMILARITY).  
 FT METAL 632 632 COPPER A2 (BY SIMILARITY).  
 FT METAL 634 634 COPPER A1 (BY SIMILARITY).  
 FT METAL 634 634 COPPER A2 (BY SIMILARITY).  
 FT METAL 638 638 COPPER A2 (BY SIMILARITY).  
 FT METAL 641 641 COPPER A1 (BY SIMILARITY).  
 SQ SEQUENCE 652 AA; 71413 MW; 40492AF4FDEDEB8 CRC64;  
 Query Match 4.4%; Score 116; DB 1; Length 652;  
 Best Local Similarity 27.8%; Pred. No. 1.2;  
 Matches 45; Conservative 18; Mismatches 63; Indels 36; Gaps 7;  
 63 DAIVTAPEVPPV---DQHPAKVYVYKMETEKVRLADGVYQFWTFGGVPGQMR 118  
 Db 505 DAIVAH-PSILSDIKSVWRNDP-----MMAETRAQAEADGVYDINMTVEVIRDNKVR 557



QY 119 -----VREGDTEVQFSNHPD-SKPHNVDFHAATGPGGAEASTFAPG 161  
 DB 558 VMSSVAPSFSESIEFTVEKEDVETVITMIDEIDDLTHGFTM-----GNVGVAMIGPQ 611  
 QY 162 HNTSFSAKLOGLVYVHCANAVPGMHINGMVGLIVEPKE 203  
 DB 612 MTSSVTFVANPGVYVYQCMFCHALHME--MGRMLVEPKE 651

## RESULT 31

NTP3\_TOBAC STANDARD; PRT; 554 AA.

AC P29162;  
 DT 01-DEC-1992 (Rel. 24, Created)  
 DT 01-DEC-1992 (Rel. 24, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE Pollen-specific protein NTP303 precursor.  
 OS Nicotiana tabacum (Common tobacco).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 OC Asteridae; lamids; Solanales; Solanaceae; Nicotiana.  
 OX NCBI\_TaxId=4097;

RN [1]  
 RP SEQUENCE FROM N.A.

RC STRAIN=cv. Petit Havana SRI; TISSUE=Pollen;  
 RX MEDLINE=92288302; PubMed=1600146;

RA Weterings K., Reijnen W., van Aarsen R., Kortstee A., Spijkers J.,  
 van Herpen M., Schrauwen J., Mullens G.,  
 "Characterization of a pollen-specific cDNA clone from Nicotiana  
 tabacum expressed during microgametogenesis and germination.";

RT Plant Mol. Biol. 18:1101-1111 (1992).  
 CC -1- FUNCTION: PRECISE FUNCTION UNKNOWN BUT PROBABLY HAS AN IMPORTANT  
 ROLE DURING GERMINATION AND/OR TUBE GROWTH.

CC -1- SUBCELLULAR LOCATION: Extracellular (Probable).  
 CC -1- TISSUE SPECIFICITY: POLLEN.

CC -1- DEVELOPMENTAL STAGE: APPEAR AFTER THE FIRST HAPLOID MITOSIS AND  
 ARE EXPRESSED DURING MICROGAMETOGENESIS, GERMINATION AND TUBE  
 GROWTH.

CC -1- INDUCTION: EXPRESSION REGULATED BY THE HAPLOID GAMETOPHYTE ITSELF.

CC -1- SIMILARITY: Contains 3 plastocyanin-like domains.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL, X61146; CAA3454.1; -  
 DR PIR: S22495; S22495.

DR InterPro: IPR001117; Cu-oxidase.  
 DR Pfam: PF00394; Cu-oxidase; 1.

KW Signal; Germination; Glycoprotein; Repeat.  
 FT SIGNAL 1 21 POTENTIAL.

FT CHAIN 22 554 POLLEN-SPECIFIC PROTEIN NTP303.  
 FT DOMAIN 22 143 PLASTOCYANIN-LIKE 1.  
 FT DOMAIN 196 286 PLASTOCYANIN-LIKE 2.  
 FT DOMAIN 411 521 PLASTOCYANIN-LIKE 3.

FT CARBOHYD 31 31 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 59 59 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 108 108 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 332 332 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 352 352 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 554 AA; 62033 MW; 9D38DAB1F52E2F85 CRC64;

Query Match 4.2%; Score 111.5; DB 1; Length 554;  
 Best Local Similarity 20.8%; Pred No. 2;

Matches 91; Conservative 43; Mismatches 160; Indels 143; Gaps 22;  
 QY 100 GVEYQFTFGGVPQGMIRVREGDTLEVQFSNHPD-----SKPHNVDFHAATGPGG 152

DB 41 GVPQGIILNGQFPBPRIKNTSSNNI VNVFNNLDEPFLFTWNGVQHRKNSWQDTPGTM 100  
 QY 153 AASAFAPGHTSTFSFKL-OPGLVYVHCANAVPGMHINGMVGLIVEPKEGLP-----K 207  
 DB 101 CP---IMPGQNTTFRQYVDQIGSYSTPTT---LHRAGGIGALNVSRRLIPFPFN 154  
 QY 208 VKEEYVVMQDFFYTKGYEEOGLQPFDEKAIREDAEVYVFNQ--SVGALTG--ENALKA 263  
 DB 155 PADENNVFVGDMYNNQ-----HKTLKK-----ILDGRTIGRPDGIINGKSA 197  
 QY 264 KVGETVR-LFVNGGPNLTSSF-----HVICEIFDKVHF 296  
 DB 198 KVGAEAKELFTWEAKTYRFRFCNLMSSVNI RFQGHMKLVELEGSHTVQNIYDSLQ 257  
 QY 297 EGGKBNHNQITLLPAGGAATTEKRVDPGVYVDNAIFPAFKGALGILKVEEENH 356  
 DB 258 HVGQ---CLSVLTAD-----QEPKQYVIV--VSSRFKQALSSVAI----- 294  
 QY 357 EYSHKQTDVAVLPFGAPQALDTPQAPKTPAPANQ-----EQIKAGATYDSNCAQH 410  
 DB 295 -----IYANGKGA--SPELP--TPPEENTEGIAMSNMQRSPFWNLTAQAR-- 339  
 QY 411 QPDGKGVNAPPPPLANSDYLNADHAPAAISYVANGLSKITVNGQYESVPAIALSDQI 470  
 DB 340 -----FN---PQGSYHVGQINIRITIKIF-NSMS-----QV 366  
 QY 471 ANVIYTLNFPKNGQ 487  
 DB 367 GGLKRYGALGISHTNGE 383

## RESULT 32

LACS\_TRAVI STANDARD; PRT; 527 AA.

AC Q99056;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Laccase 5 precursor (EC 1.10.3.2) (Benzenediol: oxygen oxidoreductase)  
 (Urishiol oxidase).

GN LCCS.

OS Trimeetes villosa (White-rot fungus).  
 OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;  
 OC Aphyllophorales; Trimeetes.

OX NCBI\_TaxId=47662;  
 RN [1]

RP SEQUENCE FROM N.A.  
 MEDLINE=97128774; PubMed=8973314;

RA Yaver D.S., Golightly E.J.;  
 RT "Cloning and characterization of three laccase genes from the  
 RT white-rot basidiomycete Trimeetes villosa: genomic organization of the  
 RT laccase gene family.";  
 RT Gene 181:95-102 (1996).  
 RN [2]

RP REVISIONS.

RA Yaver D.S., Golightly E.J.;  
 RB Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.

CC -1- FUNCTION: LIGNIN DEGRADATION AND DETOXIFICATION OF LIGNIN-DERIVED  
 CC PRODUCTS (PROBABLE).  
 CC -1- CATALYTIC ACTIVITY: 4 benzenediol + O(2) = 4 benzenesemiquinone + 2  
 CC H(2)O.

CC -1- COFACTOR: BINDS 4 CU-IONS PER MOLECULE. THREE DISTINCT CU  
 CC CENTERS KNOWN AS TYPE 1 OR BLUE, TYPE 2 OR NORMAL, AND TYPE  
 CC 3 OR COUPLED BINUCLEAR (BY SIMILARITY).  
 CC -1- SUBUNIT: Homodimer (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Secreted.

CC -1- SIMILARITY: BELONGS TO THE FAMILY OF MULTICOPPER OXIDASES.  
 CC -1- SIMILARITY: Contains 3 plastocyanin-like domains.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).)  
 CC -----  
 CC EMBL: L78078; AAB47735.2; -  
 CC PIR: JCS357; JCS357.  
 DR HSP: P37064; IAO2.  
 DR InterPro: IPR001117; Cu-oxidase.  
 DR InterPro: IPR002355; MultiCu-oxidase2.  
 DR Pfam: PF00394; Cu-oxidase; 3.  
 DR PROSITE: PS00079; MULTICOPPER\_OXIDASE1; 1.  
 DR PROSITE: PS00080; MULTICOPPER\_OXIDASE2; FALSE NEG.  
 KM Oxidoreductase; Signal; Copper; Metal-binding; Glycoprotein; Repeat;  
 KW Lignin degradation; Multigene family.  
 FT SIGNAL 1 23  
 FT CHAIN 24 527  
 FT DOMAIN 25 150  
 FT DOMAIN 162 306  
 FT DOMAIN 373 498  
 FT METAL 87 87  
 FT METAL 89 89  
 FT METAL 132 132  
 FT METAL 134 134  
 FT METAL 425 425  
 FT METAL 428 428  
 FT METAL 430 430  
 FT METAL 480 480  
 FT METAL 481 481  
 FT METAL 482 482  
 FT METAL 486 486  
 FT CARBOHYD 74 74  
 FT CARBOHYD 77 77  
 FT CARBOHYD 156 156  
 FT CARBOHYD 209 209  
 FT CARBOHYD 233 233  
 FT CARBOHYD 242 242  
 FT CARBOHYD 276 276  
 FT CARBOHYD 317 317  
 FT CARBOHYD 358 358  
 FT CARBOHYD 366 366  
 FT CARBOHYD 393 393  
 FT CARBOHYD 402 402  
 FT CARBOHYD 464 464  
 FT SEQUENCE 527 AA; 56247 MW; F1B638D56FFA478 CRC64;  
 Query Match 4.2%; Score 111; DB 1; Length 527;  
 Best Local Similarity 19.5%; Pred. No. 2.1;  
 Matches 101; Conservative 63; Mismatches 187; Indels 168; Gaps 23;  
 QY 60 PVTDAIVTAPEVPPVDRDHPAKVVKVMEVEMRLADGVEYQFTFGGQVPGQMIRV 119  
 DB 27 PVTDLTISNADVP-----DGIIRAVALAGVPPGPIITG 61  
 QY 120 REGDTLEVG-FSNHPPDSKMPHNVDF-----ATGPGGGAESAFT-----APGHTSTFS 167  
 DB 62 NKGEDEFQINIVIDLMTMETLAKSTTIHWHGFGQAKTMADG--AAFVNOGCIATGNSLYID 119  
 QY 168 FKAL-OPGLVYVCAVAPVGMHTAMNGVGLILV-----EPKEGPKYDKXEYVNO-GDFY- 220  
 DB 120 FTVPPDQAGTFWYH---SHLSTGYCDGLRGLPVYVDPDDPNASLYDVDDDTVTITLADWYH 176  
 QY 221 TKRKYGEQGLQFPDMEKALREDDAEYVFNQSVGALTGEMALKAKVGETVRLPFGNGGPNL 280  
 DB 177 TAAKLG-----PAPFAGPSVYLING-LGRFSGD-----GGGATNL 210  
 QY 281 T-----SSHFVIGEIPDKVFEKGKGNHNIOTLLPAGGAAT 318  
 DB 211 TVITVTGKKRYRRLPLVSIKDPNFTSIDGHNMTIIEVDVNHEDALVDVSIQIFAGQRY- 269  
 QY 319 TEFKVDVPGDYVL-----VDHAIFFRAF-NKGAL-----GLIKVGEENHEIYSHKOT 364  
 DB 270 -----SFLNANGSINQWIRALPNTGTTDTTGSGVASILKRYVTADIEPTTWTAT 320

QY 365 DAVLPEGAPQAIPTQZAPKTP-----APANLOE 393  
 DB 321 SVIFPLTEIDVLPDNPAPRPPQVGVGLANSIDFSFNGSNFFINNETFVPTVLL-- 378  
 QY 394 QIKAG---KATYDSCAAACHOPDGKGVNAPPEPLANSIYLANDHARAASIYANGISGKIT 450  
 DB 379 QILSGAQDAASILPFGSVYTLPSNSTIETISFPIITTDGVLNAP-----GAPHPFH 428  
 QY 451 VNGNYESVMPRIALSDQIANVITYTLNSFGNKGGLS 489  
 DB 429 LQHGF-SVRSAGSSTFNYPNVPVDRIVTSGSDNVT 466  
 RESULT 33  
 GUNC\_CELFI STANDARD; PRT; 1101 AA.  
 ID GUNC\_CELFI  
 AC P14090;  
 DT 01-JAN-1990 (Rel. 13, Created)  
 DT 01-MAR-1992 (Rel. 21, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Endoglucanase C precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase C)  
 DE (Cellulase C).  
 GN CENC.  
 OS Cellulomonas fimi.  
 OC Bacteria; Actinobacteria; Actinomycetales;  
 OC Micrococciaceae; Cellulomonadaceae; Cellulomonas.  
 OX NCBI\_Taxid=1708;  
 RN (1)  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 33-42.  
 RC STRAIN=ATCC 484;  
 RX MEDLINE=92065819; PubMed=1956299;  
 RA Coutinho J.B., Moser B., Kilburn D.G., Warren R.A.J., Miller R.C. Jr.;  
 RT "Nucleotide sequence of the endoglucanase C gene (cenc) of  
 RT Cellulomonas fimi, its high-level expression in Escherichia coli, and  
 RT characterization of its products.";  
 RL Mol. Microbiol. 5:1221-1233(1991).  
 RN (2)  
 RP SEQUENCE OF 1-64 FROM N.A., AND SEQUENCE OF 625-641.  
 RX MEDLINE=90103465; PubMed=2604391;  
 RA Moser B., Gilkes N.R., Kilburn D.G., Warren R.A.J., Miller R.C. Jr.;  
 RT "Purification and characterization of endoglucanase C of Cellulomonas  
 RT fimi, cloning of the gene, and analysis of in vivo transcripts of the  
 RT gene.";  
 RL Appl. Environ. Microbiol. 55:2480-2487(1989).  
 RN (3)  
 RP CELLULOSE-BINDING DOMAINS.  
 RX MEDLINE=92269585; PubMed=1375311;  
 RA Coutinho J.B., Gilkes N.R., Warren R.A.J., Kilburn D.G.,  
 RA Miller R.C. Jr.;  
 RT "The binding of Cellulomonas fimi endoglucanase C (Cenc) to cellulose  
 RT and Sephadex is mediated by the N-terminal repeats.";  
 RL Mol. Microbiol. 6:1243-1252(1992).  
 RN (4)  
 RP IDENTIFICATION OF IG-LIKE DOMAINS.  
 RX MEDLINE=97035265; PubMed=8880921;  
 RA Bateman A., Eddy S.R., Chothia C.;  
 RT "Members of the immunoglobulin superfamily in bacteria.";  
 RL Protein Sci. 5:1939-1942(1996).  
 RN (5)  
 RP STRUCTURE BY NMR OF 33-184.  
 RX MEDLINE=97074498; PubMed=8916925;  
 RA Johnson P.E., Joshi M.D., Tomme P., Kilburn D.G., McIntosh L.P.;  
 RT "Structure of the N-terminal cellulose-binding domain of Cellulomonas  
 RT fimi Cenc determined by nuclear magnetic resonance spectroscopy.";  
 RL Biochemistry 35:14381-14394(1996).  
 CC -1- FUNCTION: THE BIOLOGICAL CONVERSION OF CELLULOSE TO GLUCOSE  
 CC GENERALLY REQUIRES THREE TYPES OF HYDROLYTIC ENZYMES:  
 CC (1) ENDOGLUCANOLYTHOLASES THAT CUT THE CELLULOSE POLYMER CHAIN;  
 CC (2) EXOCELLULOBIOTRANSFERASES WHICH HYDROLYZE THE CELLULOSE AND OTHER  
 CC (3) BETA-1,4-GLUCOSIDASES WHICH HYDROLYZE THE CELLULOSE AND OTHER  
 CC SHORT CELLO-OLIGOSACCHARIDES TO GLUCOSE.



unusually low redox potential compared with mitochondrial cytochrome c. It is reactive with cytochrome c oxidase but not with reductases.

-----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

EMBL: AE012787; AAM71323.1; -

TIGR: CTO075; -

InterPro: IPR003088; Cyt\_C1

InterPro: IPR002323; Cyt\_C1C

InterPro: IPR000345; CytC\_heme\_bind

Pfam: PF00034; Cytochrome c; 1.

PRINTS: PR00607; CYTOCHROME C; 1.

PROSITE: PS00190; CYTOCHROME C; 1.

Electron transport; Photosynthesis; Heme; Signal; Complete proteome.

FT SIGNAL 1 22

FT CHAIN 23 108

FT BINDING 36 36

FT BINDING 39 39

FT METAL 40 40

FT METAL 82 82

SEQUENCE 108 AA; 10892 MW; 9711CB04B38825C CRC64;

-----

Query Match 4.2%; Score 109; DB 1; Length 108;

Best Local Similarity 31.0%; Pred. No. 0.37;

Matches 27; Conservative 14; Mismatches 28; Indels 18; Gaps 3;

-----

397 AGCATYDSNCAHQDQGVNAPPLNSDVLNDHRAASIVANGSKRTVNGQY 456

27 AGCATYDASCATCKTKGMMGAPRY-----GDKAMAPRIAGKMMTLVSKIKG 75

457 ---ESVMPA---IATSDQIANVITY 476

76 KGTGKMPAKGNAKLTDQVGNNAVY 102

-----

RESULT 35

CERU MOUSE

ID CERU MOUSE STANDARD; PRT; 1062 AA.

AC Q61177;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE Ceruloplasmin precursor (EC 1.16.3.1) (Ferroxidase).

GN CP.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

NCBI TaxID=10090;

OX [1]

RP SEQUENCE FROM N.A.

RA Klomp L.W.J., Fathangrazi Z.S., Choi D.W., Gitlin J.D.;

RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.

RN [2]

RP TISSUE SPECIFICITY.

RX MEDLINE=96294736; Pubmed=8690795;

RA Klomp L.W.J., Fathangrazi Z.S., Dugan L.L., Gitlin J.D.;

RL "Ceruloplasmin gene expression in the murine central nervous system.";

RL J. Clin. Invest. 98:207-215(1996).

CC -1- FUNCTION: CERULOPLASMIN IS A BLUE, COPPER-BINDING (6-7 ATOMS PER MOLECULE) GLYCOPROTEIN FOUND IN PLASMA. FOUR POSSIBLE FUNCTIONS ARE FERROXIDASE ACTIVITY, AMINE OXIDASE ACTIVITY, COPPER TRANSPORT AND HOMEOSTASIS, AND SUPEROXIDE DISMUTASE ACTIVITY.

CC -1- CATALYTIC ACTIVITY: 4 Fe(2+) + 4 H(+) + O(2) = 4 Fe(3+) + 2 H(2)O.

CC -1- COFACTOR: BINDS 6 CU-IONS PER MOLECULE. THIS PROTEIN BELONGS TO THE MULTICOPPER OXIDASES WHICH CONTAIN THREE DISTINCT CU CENTERS KNOWN AS TYPE 1 OR BLUE, TYPE 2 OR NORMAL, AND TYPE 3 OR COUPLED

BINDUCLEAR.

CC -1- TISSUE SPECIFICITY: MANY TISSUES, INCLUDING LIVER, EYE AND BRAIN.

CC -1- SIMILARITY: Contains 3 F5/8 type A domains.

-----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

EMBL: U99430; AAB07996.1; -

HSSP: P00450; 1KCV.

MGI: MGI:88476; CP.

InterPro: IPR001117; Cu-oxidase.

InterPro: IPR002355; Multic\_oxidase2.

Pfam: PF00394; Cu-oxidase; 3.

PROSITE: PS00079; MULTICOPPER OXIDASE1; 3.

PROSITE: PS00080; MULTICOPPER OXIDASE2; 1.

Oxidoreductase; Copper; Metal-binding; Glycoprotein; Plasma; Repeat;

KW Signal.

FT SIGNAL 1 19

FT CHAIN 20 1062

FT DOMAIN 20 356

FT DOMAIN 20 199

FT DOMAIN 208 356

FT DOMAIN 369 713

FT DOMAIN 369 555

FT DOMAIN 565 713

FT DOMAIN 725 1057

FT DOMAIN 725 896

FT DOMAIN 904 1057

FT DISULFID 173 199

FT DISULFID 275 356

FT DISULFID 529 555

FT DISULFID 632 713

FT DISULFID 870 896

FT METAL 120 120

FT METAL 122 122

FT METAL 179 179

FT METAL 181 181

FT METAL 990 990

FT METAL 993 993

FT METAL 995 995

FT METAL 1035 1035

FT METAL 1036 1036

FT METAL 1037 1037

FT METAL 1041 1041

FT METAL 1046 1046

FT METAL 138 138

FT CARBOHYD 226 226

FT CARBOHYD 583 583

FT CARBOHYD 625 625

FT CARBOHYD 757 757

FT CARBOHYD 922 922

SEQUENCE 1062 AA; 121159 MW; F3F52ED09A238F16 CRC64;

-----

Query Match 4.2%; Score 109; DB 1; Length 1062;

Best Local Similarity 20.5%; Pred. No. 7.1;

Matches 95; Conservative 50; Mismatches 162; Indels 156; Gaps 23;

-----

114 GGMIRAREGDTTEVQSN---HPDSKMPHNVFPHAT-----GGGGAE---ASFAPGH 162

449 GPVIMAEVDITKVFHNKGHLSIQNGVSTFANEGTYGPPASQQAASHVAPK 508

163 TSTFSF---KALQPGI-----YVHCAPVGMHIANGYGLIVREGEI-----P 206

509 TTFYEWTVKEMGFTYADVCLSKMYSAVDPT-KDIFGLGPMKICKKSGSLADGRK 567

207 KUDKEYVM-----OGDPTTKK-----YGEQG 229

568 DVDKEFLEPTVFDENESILLDDNIMFTHAPDVDEKEDFQESNKMHSNGFMWNGOS 627

```

QY 230 LQPFMEKAI-----REDABVYVNGSVGALIGENALAKKAVGETVLLFVGNG---- 276
DB 628 WPMHMGESIVWYLFSGAGNEADVHGIFYSGNTYLCKGEER-----DTANLFPHKSLTLL 681
QY 277 -GPNLTSFVHIGEIDKVFEGGKGNHNI-----QTLTPAGGAITEKVDVPGD 328
DB 682 MNDVTGTIDV--ECITTDHYTGKMKOKYVNOCORQFEDFTVYLBERTYVDVAVEMD 739
QY 329 YVLVDHAFRAFNKALGLIKVEGENHEIYSHKOTDAVLLPFGAPQALDPTOE----- 381
DB 740 Y-----SPSRAMK-----ELHHLQEQNVSNV-----LDKEFFIGSKY 774
QY 382 ---APPTAPANIQEOIKAGKATYDSNCA---ACHQPDGKGVNAPPLANSDYLNADH 434
DB 775 KKVYVQFTDSSPREQVKR--RAEDEHGLIGPPIHANVDKVVFKNATRPY----- 828
QY 435 ARAASTVANGSKITVNGQSVNPAIALSDQOLANVTYTT 477
DB 829 ---SIHAGVK-----TESSTVFTLP-----GEVATYTT 854

RESULT 36
LACS_TRAVE STANDARD; PRT; 527 AA.
AC Q12717;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Laccase 5 precursor (EC 1.10.3.2) (Benzenediol: oxygen oxidoreductase)
DE (urushiol oxidase) (diphenol oxidase) (Laccase IV).
GN LC05 OR LC04V.
OS Trametes versicolor (White-rot fungus).
OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
OC Aphyllophorales; Trametes.
OX NCBI_TaxID=5325;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=52J;
RX MEDLINE=97464057; PubMed=9322748;
RA Ong E., Pollock W.B., Smith M.;
RT "Cloning and sequence analysis of two laccase complementary DNAs from
RT the ligninolytic basidiomycete Trametes versicolor.";
RL Gene 196:113-119 (1997).
CC - FUNCTION: LIGNIN DEGRADATION AND DETOXIFICATION OF LIGNIN-DERIVED
CC PRODUCTS (PROBABLE).
CC - CATALYTIC ACTIVITY: 4 benzenediol + O(2) = 4 benzoquinone + 2
CC H(2)O.
CC - COFACTOR: BINDS 4 CU-IONS PER MOLECULE. THREE DISTINCT CU
CC CENTERS KNOWN AS TYPE 1 OR BLUE, TYPE 2 OR NORMAL, AND TYPE
CC 3 OR COUPLED BINUCLEAR (BY SIMILARITY).
CC - SUBCELLULAR LOCATION: Secreted.
CC - SIMILARITY: BELONGS TO THE FAMILY OF MULTICOPPER OXIDASES.
CC - SIMILARITY: Contains 3 plastocyanin-like domains.
CC This Swiss-Prot entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U44431; AAC94829.1; -
CC DR HSSP; P37064; IAOZ.
CC DR InterPro; IPR001117; Cu-oxidase.
CC DR InterPro; IPR002355; MultiCu_oxidase2.
CC DR Pfam; PFO0394; Cu-oxidase; 3.
CC DR PROSITE; PS00079; MULTICOPPER_OXIDASE1; 1.
CC DR PROSITE; PS00080; MULTICOPPER_OXIDASE2; FALSE_NEG.
CC Oxidoreductase; Signal; Copper; Metal-binding; Glycoprotein; Repeat;
CC Lignin degradation; Multigene family.
CC FT SIGNAL 1 23 POTENTIAL.

```

```

FT CHAIN 24 527 LACCASE 5.
FT DOMAIN 25 150 PLASTOCYANIN-LIKE 1.
FT FT 306 162 PLASTOCYANIN-LIKE 2.
FT DOMAIN 373 498 PLASTOCYANIN-LIKE 3.
FT METAL 87 87 COPPER (TYPE 2) (BY SIMILARITY).
FT METAL 89 89 COPPER (TYPE 3) (BY SIMILARITY).
FT METAL 132 132 COPPER (TYPE 3) (BY SIMILARITY).
FT METAL 134 134 COPPER (TYPE 3) (BY SIMILARITY).
FT METAL 425 425 COPPER (TYPE 1) (BY SIMILARITY).
FT METAL 428 428 COPPER (TYPE 2) (BY SIMILARITY).
FT METAL 430 430 COPPER (TYPE 3) (BY SIMILARITY).
FT METAL 480 480 COPPER (TYPE 3) (BY SIMILARITY).
FT METAL 481 481 COPPER (TYPE 1) (BY SIMILARITY).
FT METAL 482 482 COPPER (TYPE 3) (BY SIMILARITY).
FT METAL 486 486 COPPER (TYPE 1) (BY SIMILARITY).
FT CARBOHYD 74 74 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 77 77 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 156 156 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 209 209 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 233 233 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 242 242 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 276 276 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 317 317 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 358 358 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 366 366 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 393 393 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 402 402 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 464 464 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE . 527 AA; 56094 MW; D9597491F1F79825 CRC64;

```

```

Query Match 4.1%; Score 108; DB 1; Length 527;
Best Local Similarity 20.2%; Pred. No. 3.4;
Matches 104; Conservative 62; Mismatches 190; Indels 158; Gaps 26;

```

```

QY 60 PYTIDALVTHAEPVPPVBDHPRAKVYVKMETEKMRADSVYQFWTGGQVPGQMIV 119
DB 27 PYTDLTISNADVTP-----DGITRAVLNAGVFPGLITG 61
QY 120 RRGDITEVQ-FSNHPSKKPHNVDPH-----AATGGGGAASFT-----APGHSTPS 167
DB 62 NKGEFQIVNVLNLTNETYKSTTHHGIFPAQTNMADG--AAFNOCPIATGNSFLYD 119
QY 168 FRAL-OPGLVYVHCANAPVGMHIANMGV-LILVEPK---GLPKVDEYVWQ--GDYF- 220
DB 120 FTVPPQAGTFWYH-----SHLSTQYCGGLRGLVYVDDANASLYVDVDDTIVTTLADWYH 176
QY 221 TKGKYGEOGLQPFMEKAIREDAYVY-----FNGSVGALIGENALAKKAGE--TVRLF 272
DB 177 TAAKLG-----PAFPAGPDSVLLNGLGRFSGDGGATNLTVITVTKRKRYRFLV 226
QY 273 VNGGPNLTSFVHIGEIFDKVHFEGGKGNHNIQTLTPAGGAITEKVDVPG----- 327
DB 227 SISCDNFTFTS-----DGHNM--TILEVG--VNHEALVDLSQITRA 265
QY 328 ---DYVL-----VDHAFRAF-NKGAL-----GLIKVEGENHEIYSHKOTDAVLL 369
DB 266 GGRYSFTLANGSIDINWIRALPNTGTTDTGGVSAILRYPDAEIEFTNATTSVPL 325
QY 370 PFGAPQALDPTQAPKTP-----APANLQEOIKYG 398
DB 326 TETDVLPLDNPAPAPDPQVGVLDLMSLDFSFGNSFPINNETFVPPVPLV--QILSG 383
QY 399 ---KATVNSNCAACHQPDGKGVNAPPLANSDYLNADHARAASIVANGLSKIFVNGNQ 455
DB 384 AODASLLEFNGSVYTLPSNSTIEISFPLITTDGALNAP-----GAPHPFHLDGHT 433
QY 456 YESVMPAIALSDQOLANVTYTLNFGNKGGLS 489
DB 434 F-SVRSASGSSTFNANPVRADTVTSGNSGDVNT 466

RESULT 37
LACC2_AGABI

```

ID LAC2 AGABI STANDARD; PRT; 520 AA.  
 AC 012542;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DE 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE Laccase II precursor (EC 1.10.3.2) (Benzenediol: oxygen oxidoreductase)  
 DE (urithiol oxidase) (Diphenol oxidase).  
 GN LCC2.  
 OS Agaricus bisporus (Common mushroom).  
 OC Agaricales; Agaricomycota; Hymenomycetes; Homobasidiomycetes;  
 ON NCBI\_TaxID=5341;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RC STRAIN=D649;  
 RX MEDLINE=93367392; PubMed=8360614;  
 RA Perry C.R., Smith M., Britnell C.H., Wood D.A., Thurston C.F.;  
 RT "Identification of two laccase genes in the cultivated mushroom  
 Agaricus bisporus.";  
 RT J. Gen. Microbiol. 139:1209-1218(1993).  
 CC -1- FUNCTION: LIGNIN DEGRADATION AND DETOXIFICATION OF LIGNIN-DERIVED  
 PRODUCTS (PROBABLE).  
 CC -1- CATALYTIC ACTIVITY: 4 benzenediol + O(2) = 4 benzoquinone + 2  
 H(2O).  
 CC -1- COFACTOR: BINDS 4 CU-IONS PER MOLECULE. THREE DISTINCT CU  
 CENTERS KNOWN AS TYPE 1 OR BLUE, TYPE 2 OR NORMAL, AND TYPE  
 3 OR COUPLED BINUCLEAR (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- SIMILARITY: BELONGS TO THE FAMILY OF MULTICOPPER OXIDASES.  
 CC -1- SIMILARITY: Contains 3 plastocyanin-like domains.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL, L10663; AAA17035.1; -.  
 DR HSSP; P37064; IASQ.  
 DR InterPro; IPR001117; Cu-oxidase.  
 DR InterPro; IPR002355; MultiCu oxidase2.  
 DR Pfam; PF00394; Cu-oxidase\_3.  
 DR PROSITE; PS00079; MULTICOPPER\_OXIDASE1; 2.  
 DR PROSITE; PS00080; MULTICOPPER\_OXIDASE2; 1.  
 KW Oxidoreductase; Signal; Copper; Metal-binding; Lignin degradation;  
 KW Glycoprotein; Repeat; Multigene family.  
 FT SIGNAL 1 19  
 FT CHAIN 20 520 LACCASE II.  
 FT DOMAIN 21 145 PLASTOCYANIN-LIKE 1.  
 FT DOMAIN 157 305 PLASTOCYANIN-LIKE 2.  
 FT DOMAIN 375 488 PLASTOCYANIN-LIKE 3.  
 FT METAL 82 82 COPPER (TYPE 2) (BY SIMILARITY).  
 FT METAL 84 84 COPPER (TYPE 3) (BY SIMILARITY).  
 FT METAL 127 127 COPPER (TYPE 3) (BY SIMILARITY).  
 FT METAL 129 129 COPPER (TYPE 3) (BY SIMILARITY).  
 FT METAL 417 417 COPPER (TYPE 1) (BY SIMILARITY).  
 FT METAL 420 420 COPPER (TYPE 2) (BY SIMILARITY).  
 FT METAL 422 422 COPPER (TYPE 3) (BY SIMILARITY).  
 FT METAL 470 470 COPPER (TYPE 3) (BY SIMILARITY).  
 FT METAL 471 471 COPPER (TYPE 1) (BY SIMILARITY).  
 FT METAL 472 472 COPPER (TYPE 3) (BY SIMILARITY).  
 FT METAL 476 476 COPPER (TYPE 1) (BY SIMILARITY).  
 FT CARBOHYD 108 108 N-LINKED (GLCNAC... ) (POTENTIAL).  
 FT CARBOHYD 241 241 N-LINKED (GLCNAC... ) (POTENTIAL).  
 FT CARBOHYD 299 299 N-LINKED (GLCNAC... ) (POTENTIAL).  
 FT CARBOHYD 454 454 N-LINKED (GLCNAC... ) (POTENTIAL).  
 FT CARBOHYD 492 492 N-LINKED (GLCNAC... ) (POTENTIAL).  
 SQ SEQUENCE 520 AA; 57822 MW; EF10U145F0A697AA CRC64;

Query Match 4.1%; Score 107.5; DB 1; Length 520;  
 Best Local Similarity 20.5%; Pred. No. 3.6;

Matches 78; Conservative 43; Mismatches 120; Indels 139; Gaps 21;  
 QY 96 RLA-DGVEYGFMTFGGVGVGMIRREGDTIEVPSNSH-PDSKMHVDFH-----AA 146  
 DB 32 RLAPGDFEEDTVYVNGEFGTVVNGKSVRIPVNNKLTSTMRKRSYIHNGFPOART 91  
 QY 147 TGPGGAEASF--TAPGHTSFSPK-ALQPLVYHCAVAPGMHIANGVGLIV---- 199  
 DB 92 SQGDGPAPVNGCPQPPNTTFYEFVSADBSGFVWH---SHLSTQYCDGLRGAFFVYDPE 148  
 QY 200 EKEGLPKYDK-----YYVMOGDPYTKGKGEQGLQFPDMEKAIREDAEVYFNG 250  
 DB 149 DPLGHLVYVDDETVYITLAEMWYHVLAPDI--NNEFFSSGILIV-----QDS----- 192  
 QY 251 SVGALTGENALAKAKGETVRLFVNGGPRULTSSFVHIGEIFDKVPEGGKG----- 301  
 DB 193 --GLNGKGRF-----NGGPEPTP-----FAVNVGGRKRRFRVIALS 228  
 QY 302 -----ENHNIOTLLIPAGGAAITEFKVDVPGDYVLVDHAIFFAENKALGILKVEG 352  
 DB 229 CRPFPTFSVDNNL--TFMEAD-----SVEHDPVEI----- 257  
 QY 353 EENHEIYSHKQTDVAYLPEGAPOAIDT--QEAPEKT---PAPANIQEQIKAGKATYDSC 406  
 DB 258 -GNVDIYAAQVSVIL--NANQPDVNWWRAPMTGNGPDRNPINISLTIALRY---- 309  
 QY 407 AACHQPDGKGVNAPPLAN 426  
 DB 310 -----NGAPEVEPTTVN 321  
 RESULT 38  
 ID CCAA\_BACTU STANDARD; PRT; 1257 AA.  
 AC Q45754;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Pesticidal crystal protein cry12Aa (insecticidal delta-endotoxin  
 CryIIAa) (Crystalline entomocidal protoxin) (142 kDa crystal  
 protein).  
 GN CRY12AA OR CRYIIA(A) OR CRYVB.  
 OS Bacillus thuringiensis.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OX NCBI\_TaxID=1428;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 2-11.  
 RC STRAIN=NRRL B-18244 / PS33F2;  
 RA Narva K.E., Payne J.M., Schwab G.E., Hickie L.A., Galasan T.,  
 RA Sick A.J.;  
 RT "Novel Bacillus thuringiensis microbes active against nematodes, and  
 RT genes encoding novel nematode-active toxins cloned from Bacillus  
 thuringi...";  
 RL Patent number EP0462721, 27-DEC-1991.  
 CC -1- FUNCTION: ENDOTOXIN WITH NEMATOCIDAL ACTIVITY.  
 CC -1- DEVELOPMENTAL STAGE: THE CRYSTAL PROTEIN IS PRODUCED DURING  
 CC SPOULATION AND IS ACCUMULATED BOTH AS AN INCLUSION AND AS PART  
 CC OF THE SPORE COAT.  
 CC -1- MISCELLANEOUS: TOXIC SEGMENT OF THE PROTEIN IS LOCATED IN THE  
 CC N-TERMINUS.  
 CC -1- SIMILARITY: BELONGS TO THE DELTA ENDOTOXIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; L07027; AAA22355.1; -.  
 DR InterPro; IPR005638; endotoxin\_C.  
 DR InterPro; IPR005639; endotoxin\_N.

DR Pfam: PF03944; endotoxin C; 1.  
 DR Pfam: PF03945; endotoxin N; 1.  
 KW Toxin; Sporulation.  
 SQ SEQUENCE 1257 AA; 142265 MW; 3D988F6C0E3961 CRC64;

Query Match 4.1%; Score 107.5; DB 1; Length 1257;  
 Best Local Similarity 19.2%; Pred. No. 11;  
 Matches 118; Conservative 69; Mismatches 200; Indels 227; Gaps 32;

25 NOADKAQKRSSTVDAAKTANADNANSAQHOGELVDAIATVHADEPPVDRDPAPV 84  
 245 NGDKR---KSLDVNSYKNAKIKMT-----MVLDA-----LPTEDPDHYQKE 290  
 85 VVKMETEKMRADGVEYQ-----FWTEGGQVPCQMTVREGDITVEQFSNHPDS--- 135  
 291 V---EIEFRTTISPT-YQVPKKNQNTSSIVPSDLPHY-QGDLVKLEFSTRNDGL 344  
 136 -----KMP-----HNVDFMAAGPFGGAFASTAP----- 160  
 345 AKIFGTGIRTFKSPNTHETHYVDSYNTQSSGNTSRGSSNPIDLNPIISTCIRNSF 404  
 161 -----GHTSFSEFKALQPLVYVHCAPVG-----MHANGV----- 194  
 405 YKAIAGSSVUNFK--DGTQYAPAPQAPFTGAMHDSFTESDGAPEGHKLNTIYISPGDT 461  
 195 -----GLILVEPKELPKVDKEYVMQG--DYTKGKYGEQGLQPFDM 236  
 462 LKDFINVTLLSTPTINELSTKIKGF-AEKGIKNGQIMKYGKEP-INGAPVNI 519  
 237 KA-----IRDAEYV-----FNGSVGALTGNA-----LKAKVET 268  
 520 NOOTLIFEFHAKTKQYTRIRASTQYGRFLDNQGLPTSHNGYVGTGNGEN 579  
 269 VLFVFGNGGPNLTSSPHVI-----GEIFPKVHEGEGKGNHNIQTLLPAGCAITE 320  
 580 YDLYT-IGSYTTEGNTIQTQHNDRKGNVLDRIE-----VPSDIO--- 621  
 321 FKVDVPGDVVLVDHAIFFRAFNKALGILVEGSENEIYSHKQTDVAVLPEGAQADITQ 380  
 622 ---DSFQDSPEVHSTIIFDKSSPTI---WSSNHSYSHLHBSYTSQS----- 667  
 381 EAPKTPAPANLOEQIKAGKATYDSCNACHQPDGKGVNAPFPL-ANSDYLNADHARAAS 439  
 668 -----YFNHLILNL-----FHPID-----PARNHTIHVNNEDMNVYDGKDS- 703  
 440 IVANGL-----SGKIT-----VNGNOVESMPALASDOIAVITYTLNS 480  
 704 -VADGLNENKTATIPSDAWYSGTITSMHLEFNDNNFKITTPKEELSN-ELENITTOVNAL 761  
 481 FGNKGQSLADDAVA 494  
 762 FASSAQDTLASNV 775

RESULT 39  
 FLNA\_HUMAN STANDARD; PRT; 2647 AA.  
 AC P21333;  
 DT 01-MAY-1991 (Rel. 18, Created)  
 DT 01-MAY-1991 (Rel. 18, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Filamin A (Alpha-filamin) (filamin 1) (Endothelial actin-binding protein) (ABP-280) (Nonmuscle filamin).  
 GN FLNA OR FLN1 OR FLN.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OC NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.  
 RX MEDLINE=90361737; PubMed=2313161;  
 RA Gorlin J.B., Yamah R., Egan S., Stewart M., Stossel T.P., Kwiatkowski D.J., Hartwig J.H.;

RT RT "Human endothelial actin-binding protein (ABP-280, nonmuscle  
 RT RT filamin): a molecular leaf spring."  
 RT J. Cell Biol. 111:1089-1105(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96311563; PubMed=8733135;  
 RA Chen E.Y., Zollo M., Mazzarella R.A., Ciccodicola A., Chen C.-N.,  
 RA Zuo L., Heiner C., Burrough F.W., Ripetto M., Schlesinger D.,  
 RA D'Urso M.;  
 RT "Long-range sequence analysis in Xq28: thirteen known and six  
 RT candidate genes in 219.4 kb of high GC DNA between the RCP/GCP and  
 RT G6PD loci."  
 RT Hum. Mol. Genet. 5:659-668(1996).  
 RL [3]  
 RP SEQUENCE OF 1658-1772 FROM N.A.  
 RX MEDLINE=9357748; PubMed=7689010;  
 RA Maestrini E., Patroaso C., Mancini M., Rivella S., Rocchi M.,  
 RA Ripetto M., Villa A., Fratini A., Zoppe M., Vezzoni P.,  
 RA Tonolo D.;  
 RT "Mapping of two genes encoding isoforms of the actin binding protein  
 RT ABP-280, a dystrophin like protein, to Xq28 and to chromosome 7."  
 RL Hum. Mol. Genet. 2:761-766(1993).  
 RN [4]  
 RP REVIEW.  
 RX MEDLINE=21234905; PubMed=11336782;  
 RA van der Flier A., Sonnenberg A.;  
 RT "Structural and functional aspects of filamine."  
 RL Biochim. Biophys. Acta 1538:99-117(2001).  
 RN [5]  
 RP REVIEW.  
 RX MEDLINE=21146932; PubMed=11252955;  
 RA Stossel T.P., Condeelis J., Cooley L., Hartwig J.H., Noegel A.,  
 RA Schleicher M., Shapiro S.S.;  
 RT "Filamins as integrators of cell mechanics and signalling."  
 RL Nat. Rev. Mol. Cell Biol. 2:138-145(2001).  
 RN [6]  
 RP VARIANTS PH PHE-656 AND THR-1764.  
 RX MEDLINE=21423782; PubMed=11532987;  
 RA Sheen V.L., Dixon P.H., Fox J.W., Hong S.E., Kinton L., Sisodiya S.M.,  
 RA Duncan J.S., Dubau F., Scheffer I.E., Schachter S.C., Walner A.,  
 RA Henry J., Cirino P., Kanuro K., Dimario F., Berg M., Kuznietchy R.,  
 RA Cole A.J., Bromfield E., Biber M., Schomer D., Whales J., Silver K.,  
 RA Mochida G.H., Berkovic S.F., Andermann F., Andermann E., Dobyns W.B.,  
 RA Wood N.W., Walsh C.A.;  
 RT "Mutations in the X-linked filamin 1 gene cause periventricular  
 RT nodular heterotopia in males as well as in females."  
 RL Hum. Mol. Genet. 10:1775-1783(2001).  
 CC [1]- FUNCTION: Promotes orthogonal branching of actin filaments and  
 CC links actin filaments to membrane glycoproteins. Anchors various  
 CC transmembrane proteins to the actin cytoskeleton and serves as a  
 CC scaffold for a wide range of cytoplasmic signaling proteins.  
 CC [2]- SUBUNIT: Homodimer. Interacts with cVsp, interacts with various  
 CC other binding partners in addition to filamentous actin.  
 CC [3]- SUBCELLULAR LOCATION: PERIPHERAL CYTOPLASM.  
 CC [4]- TISSUE SPECIFICITY: Ubiquitous.  
 CC [5]- PTM: PHOSPHORYLATION EXTENT CHANGES IN RESPONSE TO CELL  
 CC ACTIVATION.  
 CC [6]- PTM: The N-terminus is blocked.  
 CC [7]- DISEASE: Defects in FLNA are the cause of periventricular  
 CC heterotopia (PH) also called nodular heterotopia, bilateral  
 CC periventricular (NHP or BPNH). PH is an X-linked developmental  
 CC dominant disorder in which many neurons fail to migrate into the  
 CC cerebral cortex. They remain as nodules lining the ventricular  
 CC surface. In heterozygous females these neurons presumably  
 CC represent those cells that, after X-chromosome inactivation,  
 CC contain the active X chromosome with the filamin mutation. Most  
 CC heterozygous affected males die early during embryogenesis, whereas  
 CC heterozygous females have normal intelligence but suffer from  
 CC seizures and various manifestations outside the central nervous  
 CC system, especially related to the vascular system. This implies  
 CC that essential embryonic cell migration can only occur in FLNA-  
 CC expressing cells.  
 CC [8]- SIMILARITY: THE ACTIN-BINDING DOMAIN IS OF A TYPE FOUND IN MANY

CC ACTIN-BINDING PROTEINS (SUCH AS ACTININ, DYSTROPHIN, FIMBRIN,  
 CC ABP-120, ABP-180, OR BETA-FODRIN).  
 CC -1- SIMILARITY: Contains 2 calponin-homology (CH) domains.  
 CC -1- SIMILARITY: Contains 24 filament repeats.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: X53416; CAA37495.1; -  
 CC EMBL: U44140; AAA2644.1; -  
 CC EMBL: X70082; CAA49687.1; -  
 CC EMBL: X70085; CAA49690.1; -  
 CC PIR: A37098; A37098.  
 CC HSSP: P13466; 1KSR.  
 CC Gene: HGNC:3754; FLNA.  
 CC MIM: 300017; -  
 CC MIM: 300049; -  
 CC GO: GO:0015629; C:actin cytoskeleton; TAS.  
 CC GO: GO:0003780; P:actin cross-linking activity; TAS.  
 CC GO: GO:0007012; P:actin cytoskeleton reorganization; TAS.  
 CC GO: GO:0007399; P:neurogenesis; TAS.  
 CC InterPro: IPR001589; Actbind\_actin.  
 CC InterPro: IPR001298; Calponin-like.  
 CC InterPro: IPR00307; CH: 2.  
 CC Pfam: PF00630; Filamin; 24.  
 CC SMART: SM00033; CH: 2.  
 CC SMART: SM00557; IG\_FLMN; 24.  
 CC PROSITE: PS00019; ACTININ\_1; 1.  
 CC PROSITE: PS00020; ACTININ\_2; 1.  
 CC PROSITE: PS50021; CH: 2.  
 CC PROSITE: PS50194; FILAMIN\_REPEAT; 24.  
 CC Actin-binding; Phosphorylation; Repeat; Polymorphism;  
 CC disease mutation.  
 KW  
 FT DOMAIN 1 274 ACTIN-BINDING (HEAD).  
 FT DOMAIN 43 149 CH 1.  
 FT DOMAIN 166 266 CH 2.  
 FT REPEAT 276 374 FILAMIN 1.  
 FT REPEAT 376 474 FILAMIN 2.  
 FT REPEAT 475 570 FILAMIN 3.  
 FT REPEAT 571 663 FILAMIN 4.  
 FT REPEAT 667 763 FILAMIN 5.  
 FT REPEAT 764 866 FILAMIN 6.  
 FT REPEAT 867 965 FILAMIN 7.  
 FT REPEAT 966 1061 FILAMIN 8.  
 FT REPEAT 1062 1154 FILAMIN 9.  
 FT REPEAT 1155 1249 FILAMIN 10.  
 FT REPEAT 1250 1349 FILAMIN 11.  
 FT REPEAT 1350 1442 FILAMIN 12.  
 FT REPEAT 1443 1539 FILAMIN 13.  
 FT REPEAT 1540 1636 FILAMIN 14.  
 FT REPEAT 1649 1740 FILAMIN 15.  
 FT REPEAT 1741 1778 HINGE 1.  
 FT REPEAT 1779 1860 FILAMIN 16.  
 FT REPEAT 1861 1950 FILAMIN 17.  
 FT REPEAT 1951 2039 FILAMIN 18.  
 FT REPEAT 2042 2131 FILAMIN 19.  
 FT REPEAT 2132 2230 FILAMIN 20.  
 FT REPEAT 2233 2325 FILAMIN 21.  
 FT REPEAT 2327 2420 FILAMIN 22.  
 FT REPEAT 2424 2516 FILAMIN 23.  
 FT REPEAT 2517 2551 HINGE 2.  
 FT REPEAT 2552 2646 FILAMIN 24.  
 FT REPEAT 2647 2664 SELF-ASSOCIATION SITE, TAIL.  
 FT SITE 1761 1762 CLEAVAGE (BY CALPAIN).  
 FT VARIANT 320 320 V -> A (IN dbSNP:1064816).  
 FT VARIANT 370 370 F -> L (IN dbSNP:1064817).

FT FT /FtId=VAR 012832.  
 FT VARIANT 552 552 V -> A (IN dbSNP:730319).  
 FT VARIANT 656 656 /FtId=VAR 012833.  
 FT VARIANT 656 656 L -> F (IN PH).  
 FT VARIANT 1764 1764 /FtId=VAR 012834.  
 FT VARIANT 1764 1764 A -> T (IN PH).  
 FT CONFLICT 1772 1772 /FtId=VAR 012835.  
 FT CONFLICT 2634 2634 A -> G (IN REF. 3).  
 FT CONFLICT 2634 2634 H -> D (IN REF. 2).  
 SQ SEQUENCE 2647 AA; 280759 MW; 6C1A07041DFA3D42 CRC64;  
 Query Match 4.1%; Score 107.5; DB 1; Length 2647;  
 Best Local Similarity 20.8%; Pred. No. 30;  
 Matches 115; Conservative 71; Mismatches 217; Indels 149; Gaps 30;  
 QY 4 PTLITITICALSALMLSGCNSQADKAQPSSTYDAAKTANMDNMSQHQSLPIYD 63  
 DB 2151 PSVANVGSCHDLS-LKPEISIQ-DMTAQVTS--PSGKTHEAIVEGENTTYCIRFPV 2204  
 QY 64 AIV-----THAPEVP-----PVDDRHPRKVVVKKMETVEKVRADGV--EYOF 105  
 DB 2205 AEMGHTIVSVKXKQGVHVGSPQFTVGPLGEGAKVAGRGLE--RAEGVAFESI 2261  
 QY 106 WTFGGQVPGQMIKVRREGDTIEVQFSNHPDS-----KMPHNVFHAATPGGGAENSF 157  
 DB 2262 WTRBAGAGGLAIVGSPKAEISFEDRKDGSQGVAVVQEPDQYEVSVKFNHEHIDSPF 2321  
 QY 158 TAP-----GHSTSEFKALQ-PGLVYVYCAVAPVGMHIANGYGILVEPKGLPKVDKE 211  
 DB 2322 VVPVAPSPGDARRLTVSSILQESGLKVNQAPSPAVSLNKAIGALIDAKVHSPGAL---EE 2377  
 QY 212 YVWQGDPTKGYKGYGEOGLQPFDEKALREDAEYV--FNGSVGLTGENALKAKVGET 268  
 DB 2378 CYVTEID--QDKAVRPIP-----RENVYLIDVFNQT--HIG-SPFKRVGEP 2423  
 QY 269 VRLFVNGG-PNLTVSSFVHIGEIFDKVHEGKGEN--HNIGTLIPAGAAIT----- 319  
 DB 2424 -----GHGDDPLVSAV-----GAGLEGVTNPAPFVNTSNAGALSVTIDGPS 2470  
 QY 320 EKKVD-----VPGDYVL-----VDHAFPRANKKALGILKVEGE-- 353  
 DB 2471 KYKMDCOBCEPREGVRYTTPMAPGSYLISIKYGGPHIGSSPKA-----KVTGPR 2521  
 QY 354 -ENHEIYSHKOTDAYVLEPGAPQALDIOEAPKTPAPMLQEOIKAG-----KATY 402  
 DB 2522 VSNHSLH--ETSSVFDLSLTATCAPQHGAPGPPADASKVYVAGGLGSKAVYQKSSF 2578  
 QY 403 DSNCA-----ACHOP-----DGKGVNAPRPPLANSDYLNADHARAASIVANGLSGKI 449  
 DB 2579 TVDCSKAGNNMLVGVHGRTPCETILVGVGSRLYSVYLKDKGEYTLVVK--WGHE 2635  
 QY 450 TVNGNOYESVMP 461  
 DB 2636 HIRGSPYRVVP 2647  
 RESULT 40  
 OMPB\_RICRI STANDARD; PRT; 1654 AA.  
 AC O53047;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Outer membrane protein B precursor (168 kDa surface-layer protein)  
 DE (surface protein antigen) (Cell surface antigen 5) (Scas) (rOmpB)  
 DE (rOmp B) (contains: 120 kDa surface-exposed protein (surface protein  
 antigen) (120 kDa outer membrane protein ompB); 32 kDa beta peptide).  
 GN OMPB  
 OS Rickettsia rickettsii.  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;  
 OC Rickettsiaceae; Rickettsia.  
 OX NCBI\_TaxId=783;  
 RN [1]



RP SEQUENCE FROM N.A.  
 RC STRAIN=R;  
 RX MEDLINE=92167802; PubMed=1724278; Pollicastro P.F., Hackstadt T.;  
 RA Gilmore R.D. Jr., Cieplik W. Jr., "The 120 kDa surface-exposed protein of Rickettsia  
 RT rickettsii is encoded by an unusually long open reading frame:  
 RL evidence for protein processing from a large precursor.";  
 RL Mol. Microbiol. 5:2361-2370(1991).  
 RP SEQUENCE OF 279-1654 FROM N.A.  
 RC STRAIN=R;  
 RX MEDLINE=90136087; PubMed=2515418;  
 RA Gilmore R.D. Jr., Joeste N., McDonald G.A.;  
 RT "Cloning, expression and sequence analysis of the gene encoding the  
 RT 120 kD surface-exposed protein of Rickettsia rickettsii.";  
 RL Mol. Microbiol. 3:1579-1586(1989).  
 CC -1- FUNCTION: THE 120 kDa SURFACE-EXPOSED PROTEIN IS A MAJOR  
 CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL  
 CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.  
 CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-  
 CC LAYER WITH HEXAGONAL SYMMETRY.  
 CC -1- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPA FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; X16353; CAA34403.1; -  
 DR PIR; S18227; S18227.  
 DR InterPro; IPR006315; Autotransport.  
 DR InterPro; IPR005546; Autotransporter.  
 DR Pfam; PF03797; Autotransporter; 1.  
 DR TIGRPFAMs; TIGR01414; autotrans\_bar1; 2.  
 DR Antigen; S-layer; Cell wall.  
 FT CHAIN 1 1333 120 kDa SURFACE-EXPOSED PROTEIN.  
 FT CHAIN 1334 1654 32 kDa BETA PEPTIDE.  
 FT DOMAIN 1181 1188 POLY-THR.  
 FT SEQUENCE 1654 AA; 168184 MW; D7AB70FB7087F618 CRC64;

Query Match 4.1%; Score 107; DB 1; Length 1654;  
 Best Local Similarity 20.9%; Pred. No. 18;  
 Matches 128; Conservative 65; Mismatches 228; Indels 190; Gaps 29;

QY 6 LIKTLILICALMSSGCSNOADKAAQPKSIVDAAKT--ANADNAASQEHGELPVI 62  
 DB 11 LISAGLVSTATIVASFGASAMGAIQONRTTNGAATVVDGAFDQTAAPAVG--VAL 68  
 QY 63 DAIVTHAPEVPPVDRDHPA-----KVVVKMETVEKMKRLADGVEYQFWTF 108  
 DB 69 NNAVIT--ANANGININPTPAGSPNGILLTANNLAATVASEDT--TLGFTTVVNHANSP 123  
 QY 109 G-----GOVPGQMI RVEGDTIEVOFSN-----HPDSKMPHNVDFHAATGP 149  
 DB 124 NLTLAGKTLITTGCGVTNAQAATKNAQNVVQFNNGAIDNNDKVGRIIDF----- 177  
 QY 150 GGGAEASFTAPGHTSTFSFKALQPLGLVYHCAVAPV---GMHIANGMVGLILVEPKGL 205  
 DB 178 --GAPD-----STLVFNLANP-----TTQKAPLILGDNAVIANGVNGTLNV----- 216  
 QY 206 PKVDKEYVYMGQDFYTKGKYGEGGLQPFDEKAIR--EDAEVYVNGSVGALTGENALKAK 264  
 DB 217 -----TNG--FTIOVSNKSFATVAKAINIADGGIIPNTDA--NNANTLNLQ 257  
 QY 265 VGEYVALFVNGGPN-----LTSSEPHVIGEI-----FDKVFEEGKGGENHNI 306  
 DB 258 AGGTTINFTGTGTGLVLLSKHAATNTNITGSLGKLVIEFTVAVDGQLTANAGA 317  
 QY 307 QTTLLI---PAGGAATFERKVDVPGDYVLVDHAIFFAENKGAIGLIKVEGE---ENHEIY 359

DB 318 ANAVIGTNNAGRAA-----GFVVSVDNG-----KVATIDGOVYAKDMVIO 358  
 QY 360 SHKQTDVAVYLPFGAPQAIPTQAPKTPAP-----ANLQEQIK----- 396  
 DB 359 SANATGOVVFRIIVDVGADGTTAFRTASKVITTDQSNFGNDFGLAAQIKVPAITLT 418  
 QY 397 ---AGKATYDSNCAACHQPDGKV---PNAFPPLANSDYLANDHAPPAASIV----- 441  
 DB 419 GNFTGDASNPGNTAGVITFDANGTLESASADANVAVTNNITALEASGAGVQLSGTHAAE 478  
 QY 442 -ANGUSGT-----TVNG--NOYESVVPALALSDQOLANVITTTLNSFGKNGGGLSAD 491  
 DB 479 LRLGNAGSIFKLADGTVINGKVNQOTALVGALAAAGTTLDGSAITTT-GDIGNAGGAAALQ 537  
 QY 492 DVA---KAKKT 499  
 DB 538 RITLANDAKKT 548

Search completed: August 27, 2003, 18:30:06  
 Job time : 30 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 27, 2003, 18:33:27 ; Search time 26 Seconds

(Without alignments)  
907.977 Million cell updates/sec

Title: US-10-088-045-2

Perfect score: 502  
Sequence: 1 MSKPTLIKTLICALSALML.....NKGQLSADDVAKAKTKPEN 502

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 127863 seqs, 47026705 residues

Word size : 0

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database : SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	34	6.8	392	1 ANIA_NERGO	Q02219 neisseria g
2	12	2.4	330	1 NIR_ALCKX	P81445 alcaligenes
3	12	2.4	363	1 NIR_PSECL	Q06006 pseudomonas
4	10	2.0	377	1 NIR_RHINE	O60214 rhizobium h
5	8	1.6	134	1 RCSF_ECOLI	P28653 escherichia
6	8	1.6	389	1 PORB_MERJA	O57715 methanococc
7	8	1.6	432	1 TIG_THERN	O8rc26 thermoaer
8	8	1.6	518	1 SAP_CHICK	O13035 gallus gall
9	8	1.6	604	1 YFIC_BACSU	P54719 bacillus su
10	7	1.4	86	1 C555_CHLLE	P00123 chlorobium
11	7	1.4	108	1 C555_CHLLE	O8K953 chlorobium
12	7	1.4	109	1 RBS_PROHO	P27569 prochloroth
13	7	1.4	111	1 RBS_SYNR2	O44178 synechococc
14	7	1.4	111	1 YC39_AQUAE	O67284 aquifex aeo
15	7	1.4	130	1 RS9_LACIA	O9cdg7 lactococcus
16	7	1.4	157	1 GREB_BRUME	O9ydg6 bruceella me
17	7	1.4	172	1 KCH5_MOUSE	O92063 mus musculu
18	7	1.4	187	1 EFAS_MOUSE	O08545 mus musculu
19	7	1.4	193	1 YEAY_ECOLI	P76255 escherichia
20	7	1.4	194	1 RISE_ARCFU	O27965 archaeoglob
21	7	1.4	212	1 RPSD_MYCTU	O50712 mycobacteri
22	7	1.4	218	1 YIDX_ECOLI	P14611 escherichia
23	7	1.4	226	1 TPIS_METBR	O74025 methanobact
24	7	1.4	230	1 YX09_CAEEL	O11115 caenorhabdi
25	7	1.4	238	1 EFAS_HUMAN	P52737 homo sapien
26	7	1.4	239	1 YGIP_YEAST	P53223 saccharomyc
27	7	1.4	272	1 FMDC_METMP	O31112 methanococc
28	7	1.4	319	1 SYGA_COXBU	P94616 coxiella bu
29	7	1.4	328	1 HAM1_STRP3	O8K847 streptococc
30	7	1.4	328	1 HAM1_STRP8	O8P242 streptococc
31	7	1.4	328	1 HAM1_STRPY	O8A186 streptococc
32	7	1.4	351	1 PYRB_LACIE	O60252 lactobacilli
33	7	1.4	352	1 AVRC_PSSSG	P13836 pseudomonas

34	7	1.4	374	1 METL_PEA	P49613 pisum sativ
35	7	1.4	374	1 NIR_RHOSH	Q53339 rhodobacter
36	7	1.4	376	1 NIR_ALCEA	P38501 alcaligenes
37	7	1.4	378	1 NIR_ACHY	P25006 achromobact
38	7	1.4	378	1 YMA2_MYCTU	O10400 mycobacteri
39	7	1.4	387	1 RECC_LACIA	O01840 lactococcus
40	7	1.4	393	1 METK_CATRO	O96551 catarranthu
41	7	1.4	393	1 METK_PINBN	P50300 pinus banks
42	7	1.4	393	1 METL_LYCES	P43281 lycopersico
43	7	1.4	404	1 KAPR_COLTR	O42794 colletoeric
44	7	1.4	436	1 GDPE_BOVIN	P55469 bos taurus
45	7	1.4	444	1 TIG_RHOCA	O68129 rhodobacter
46	7	1.4	459	1 YGW9_YEAST	P53083 saccharomyc
47	7	1.4	462	1 LEU2_LISIN	O08580 mus musculu
48	7	1.4	462	1 LEU2_LISIN	O92a26 listeria in
49	7	1.4	502	1 ATPA_BACHD	O10087 schizosacch
50	7	1.4	508	1 RPA2_THEVO	O8txb7 methanopyru
51	7	1.4	512	1 FUS1_YEAST	P55469 rhizobium s
52	7	1.4	519	1 ERRI_HUMAN	O8u87 agrobacteri
53	7	1.4	525	1 AROF_ARATH	O9h25 p glucosami
54	7	1.4	550	1 YAO8_SCHPO	P52326 pseudomonas
55	7	1.4	571	1 SYE_METKA	P38822 saccharomyc
56	7	1.4	586	1 YAGM_RHISN	P15245 trichosporo
57	7	1.4	595	1 SYD_AGRIS	O90625 gallus gall
58	7	1.4	610	1 GLMS_PSEAE	O05065 streptomyc
59	7	1.4	615	1 RPSD_PSEFL	P49739 xenopus lae
60	7	1.4	633	1 BZ21_YEAST	O08269 saccharomyc
61	7	1.4	664	1 PH2M_TRICU	O63472 rattus norv
62	7	1.4	706	1 Z151_CHICK	O18965 bos taurus
63	7	1.4	733	1 MUTB_STRCM	O8ncm2 homo sapien
64	7	1.4	807	1 MCW3_XENLA	O9ep3 rattus norv
65	7	1.4	859	1 ALRI_YEAST	O95259 homo sapien
66	7	1.4	962	1 KCHI_RAT	O60603 mus musculu
67	7	1.4	987	1 KCHI_BOVIN	P48053 caenorhabdi
68	7	1.4	988	1 KCH5_HUMAN	P33760 saccharomyc
69	7	1.4	988	1 KCH5_RAT	O02280 drosophila
70	7	1.4	989	1 KCHI_HUMAN	P18280 strongyloce
71	7	1.4	989	1 KCHI_MOUSE	P47310 mycoplasma
72	7	1.4	995	1 YPDI_CAEEL	P21519 drosophila
73	7	1.4	1030	1 PEK6_YEAST	P87319 schizosacch
74	7	1.4	1174	1 CIKE_DROME	P18479 z genome po
75	7	1.4	1174	1 KCRP_STRPU	O89330 z genome po
76	7	1.4	1331	1 Y064_MYCGE	O34367 myoxocephal
77	7	1.4	1596	1 MAM_DROME	P20421 myoxocephal
78	7	1.4	3011	1 YOS1_SCHPO	P35808 schistocerc
79	7	1.4	3080	1 POLG_ZYMVC	O927369 sulfolobus
80	7	1.4	3083	1 POLG_ZYMVC	O33523 rickettsia
81	7	1.4	3083	1 ANP3_ZYMVS	P32150 escherichia
82	7	1.2	33	1 ANP3_MYOSC	O33508 rickettsia
83	7	1.2	33	1 ANP5_MYOGA	O33543 rickettsia
84	7	1.2	33	1 AKH2_SCHGR	O33545 rickettsia
85	7	1.2	39	1 RS27_SULTO	O33544 rickettsia
86	7	1.2	72	1 FMT_RICEP	O33575 rickettsia
87	7	1.2	72	1 YIIF_ECOLI	P38393 escherichia
88	7	1.2	73	1 FMT_RICPA	P00107 pavlova lut
89	7	1.2	73	1 FMT_RICRH	P00111 plectonema
90	7	1.2	73	1 FMT_RICRI	P00111 porphyra te
91	7	1.2	73	1 FMT_RICSI	P00110 bunilleriop
92	7	1.2	73	1 KIL_ECOLI	P00114 synechococc
93	7	1.2	83	1 CYC6_PAVLU	P00117 plectonema
94	7	1.2	83	1 CYC6_PLEBO	P00111 porphyra te
95	7	1.2	85	1 CYC6_PORTE	P00114 synechococc
96	7	1.2	85	1 CYC6_BUMFI	P00114 synechococc
97	7	1.2	87	1 CYC6_STNLI	P00114 synechococc
98	7	1.2	87	1 CYC6_STNLI	P00114 synechococc
99	7	1.2	87	1 CYC6_STNLI	P00114 synechococc
100	7	1.2	88	1 CYC6_BRYMA	P14448 bryopsis ma
101	7	1.2	88	1 Y073_BACHD	O9K518 bacillus ha
102	7	1.2	89	1 CYC6_MONBR	O90909 monoraphidi
103	7	1.2	89	1 CYC6_SCEOB	P57736 scenedesmus
104	7	1.2	89	1 CYC6_SPIMA	P00118 spirulina m
105	7	1.2	91	1 CYC6_CUAGO	P83391 cladophora
106	7	1.2	94	1 NIBW_NEUCR	O07842 neurospora

107	6	1.2	95	1	RR20_CVAPA	P48140	Cyanophora	180	6	1.2	149	1	FUR_VIBPA	O24755	vibrio para
108	6	1.2	97	1	CH10_BUCGU	Q9f44f	buchnera ap	181	6	1.2	149	1	FUR_VIBPA	P33117	vibrio vuln
109	6	1.2	97	1	GATC_LISIN	Q92a62	listeria in	182	6	1.2	150	1	FUR_VIBCH	P33087	vibrio chol
110	6	1.2	97	1	GATC_LISMO	P58817	listeria mo	183	6	1.2	150	1	VC18_VACC	P31102	vaccinia vi
111	6	1.2	98	1	FER3_ANAVA	P46050	anabaena va	184	6	1.2	150	1	YM22_VIBVU	O8d8t7	vibrio vuln
112	6	1.2	104	1	RL27_MYGE	P47476	mycoplasma	185	6	1.2	150	1	YPSL_STYEN	P25508	synecococ
113	6	1.2	105	1	NAL2_CANAL	P46594	candida alb	186	6	1.2	151	1	YF13_ECOLI	Q93942	escherichia
114	6	1.2	108	1	NIRD_ECOLI	P23675	escherichia	187	6	1.2	156	1	YXIF_BACSU	P42298	bacillus su
115	6	1.2	108	1	NIRD_SALTU	P40769	salmonella	188	6	1.2	158	1	RS9_BRUME	O8y9j0	bruceella me
116	6	1.2	109	1	CYC6_CVACA	Q9t1v1	cyanidium c	189	6	1.2	161	1	HBPL_PARAD	O43496	parasponia
117	6	1.2	109	1	RLA1_MAIZE	P52855	zea mays (m	190	6	1.2	162	1	Y875_MYCTU	O10537	mycobacteri
118	6	1.2	109	1	YM33_MAIZE	O82106	zea mays (m	191	6	1.2	162	1	Y9H4_PSEAE	Q9h110	pseudomonas
119	6	1.2	110	1	CYC6_PORPU	P51200	porphyra pu	192	6	1.2	163	1	C550_SYNEL	P56150	synecococ
120	6	1.2	110	1	CYC6_PORYE	Q8wk18	porphyra ye	193	6	1.2	163	1	H1_TETTH	P10156	tetrahymena
121	6	1.2	111	1	CYC6_SYNP7	P25935	synecococ	194	6	1.2	164	1	GREB_MYCLE	P46080	mycobacteri
122	6	1.2	112	1	CYC6_SYNEL	P56534	synecococ	195	6	1.2	164	1	GREB_MYCTU	O53428	mycobacteri
123	6	1.2	112	1	CYC6_SYNVU	Q9f119	synecococ	196	6	1.2	164	1	YOH4_ECOL6	O8f416	escherichia
124	6	1.2	112	1	RLA1_DROME	P08570	drosophila	197	6	1.2	164	1	YOH4_ECOLI	P52082	escherichia
125	6	1.2	112	1	RLA2_BABBO	P27055	babesia bov	198	6	1.2	164	1	YOH4_SALTI	Q823q8	salmonella
126	6	1.2	113	1	YG35_BPMV4	Q04775	lactococcus	199	6	1.2	164	1	YOH4_SALTY	O8m2q8	salmonella
127	6	1.2	114	1	MAUL_METEX	Q49129	methylobact	200	6	1.2	166	1	PLAS_PRIAG	O22646	fructillaria
128	6	1.2	114	1	RLA1_CHICK	P18660	gallus gall	201	6	1.2	166	1	RNH2_LACLC	O30415	lactococcus
129	6	1.2	114	1	RLA1_MOUSE	P47955	mus musculus	202	6	1.2	167	1	YDER_ECOLI	P77294	escherichia
130	6	1.2	115	1	IF1A_PYRFU	Q8u0k5	pyrococcus	203	6	1.2	169	1	HXA9_CHICK	Q98924	gallus gall
131	6	1.2	115	1	IF1A_PYRFU	O59280	pyrococcus	204	6	1.2	174	1	*PAP3_MOUSE	O09049	mus musculus
132	6	1.2	115	1	RL19_BUCAI	P57477	buchnera ap	205	6	1.2	174	1	POPB_RALSO	Q981	raletonia s
133	6	1.2	115	1	RNPA_BUCAI	P57130	buchnera ap	206	6	1.2	175	1	AXIK_ARATH	O24410	arabidopsis
134	6	1.2	117	1	CYC6_SYNP2	O30881	synecococ	207	6	1.2	176	1	UCRI_BRAJA	P51130	bradyrhizob
135	6	1.2	117	1	RL18_HAEIN	P44356	haemophilus	208	6	1.2	180	1	CTGI_HUMAN	P78358	homo sapien
136	6	1.2	117	1	RL18_VIBCH	Q9kx00	vibrio chol	209	6	1.2	181	1	Y573_RICPR	Q92x88	rickettsia
137	6	1.2	117	1	RL18_VIBPR	P52863	vibrio proc	210	6	1.2	182	1	IDI_BREIN	Q9Kf75	bruyabacter
138	6	1.2	118	1	YG35_BPLH	Q04769	lactococcus	211	6	1.2	183	1	NUSM_PEA	O35639	pisum sativ
139	6	1.2	120	1	CYC6_SYNP3	P46445	caenorhabdit	212	6	1.2	184	1	RRF_AQUAE	O66328	aquifex aeo
140	6	1.2	120	1	MIFH_CABEL	O18785	caenorhabdit	213	6	1.2	186	1	ATPD_RHOBL	O50425	methylobact
141	6	1.2	122	1	LITH_PIG	Q29191	sus scrofa	214	6	1.2	186	1	DHML_METFL	P06691	pseudomonas
142	6	1.2	122	1	Y590_VIBVU	Q8de3h	vibrio vuln	215	6	1.2	186	1	TNR5_PSEAE	P06692	escherichia
143	6	1.2	124	1	PA21_BOTJA	P81243	botrytis ja	216	6	1.2	186	1	TNR7_ECOLI	O59453	methylobact
144	6	1.2	124	1	RL7_RALSO	Q8uxz7	raletonia s	217	6	1.2	187	1	DHML_METME	O811h5	methylobact
145	6	1.2	124	1	RL17_XANCP	Q93266	xanthomonas	218	6	1.2	187	1	RRF_PARZE	P37194	paracoccus
146	6	1.2	129	1	ACP2_HORVU	P08817	hordeum vul	219	6	1.2	188	1	SLP_ECOLI	P27882	escherichia
147	6	1.2	129	1	ACP2_PSEAE	O52761	hordeum vul	220	6	1.2	189	1	ERV1_YEAST	P27882	saccharomyc
148	6	1.2	129	1	RL17_PSEAE	Q978n4	streptococ	221	6	1.2	189	1	GRPE_HELPY	P55970	helicobacte
149	6	1.2	130	1	RS9_STRPN	Q99y08	streptococ	222	6	1.2	191	1	GRPE_HELPY	Q92mw3	helicobacte
150	6	1.2	131	1	HS1R_HAEIN	P44754	haemophilus	223	6	1.2	193	1	RNH2_RICCN	O92105	rickettsia
151	6	1.2	131	1	RL32_CANAL	Q94008	candida alb	224	6	1.2	194	1	CLMP_CAMJE	P54413	campylobact
152	6	1.2	132	1	RS9_MYCEN	P57179	mycoplasma	225	6	1.2	194	1	HSB_XENLA	P22845	xenopus lae
153	6	1.2	134	1	ACP1_BRANA	P10352	brassica na	226	6	1.2	195	1	NUGM_MARRO	P34944	maricantia
154	6	1.2	134	1	ACP2_BRANA	P17650	brassica na	227	6	1.2	195	1	YLM2_CABEL	P34376	caenorhabdit
155	6	1.2	134	1	ACP5_BRANA	P08971	brassica na	228	6	1.2	196	1	HS4_XENLA	P22844	xenopus lae
156	6	1.2	134	1	ACP BRAM	P07088	brassica ca	229	6	1.2	198	1	NODA_RHIME	P03266	thizobium m
157	6	1.2	136	1	ACP1_CASGL	P93092	casuarina g	230	6	1.2	198	1	VCO7_ADE02	P03266	human adeno
158	6	1.2	136	1	RL28_BAT	P17702	rattus norv	231	6	1.2	199	1	UREG_HELPY	Q92mz7	human adeno
159	6	1.2	139	1	PSAD_ANASP	P58573	anabaena sp	232	6	1.2	199	1	UREG_HELPY	Q92mz7	human adeno
160	6	1.2	140	1	YMB4_PYRAE	Q82ct1	pyrobaculum	233	6	1.2	200	1	COWA_BRAJA	P30963	bradyrhizob
161	6	1.2	141	1	PAB2_FUSNN	O87630	fusobacteri	234	6	1.2	202	1	Y357_CAMJE	Q9p1e6	campylobact
162	6	1.2	141	1	YEDD_SALTU	O27129	methanobact	235	6	1.2	203	1	YCXW_PORPU	P51360	porphyra pu
163	6	1.2	142	1	YEDD_SALTU	O06339	salmonella	236	6	1.2	204	1	KADA_AERPE	Q9ydd2	aeropyrum p
164	6	1.2	142	1	RLN1_MYCTU	P96931	mycobacteri	237	6	1.2	204	1	UREG_BACSB	O07403	bacillus sp
165	6	1.2	142	1	RNPA_XYLA	Q9p9u0	xyella fae	238	6	1.2	204	1	UREG_STRAX	P42877	staphylococ
166	6	1.2	142	1	RS12_THEAC	Q9hly2	thermoplasma	239	6	1.2	204	1	UREG_STRSL	Q55057	streptococ
167	6	1.2	142	1	RS12_THEVO	Q97cd8	thermoplasma	240	6	1.2	205	1	RUVA_AGRT5	Q8u5k5	agrobacteri
168	6	1.2	142	1	YE63_METTM	O50770	methanobact	241	6	1.2	205	1	UREG_ECOLI	P03387	escherichia
169	6	1.2	144	1	AMSI_ERWAM	Q46630	erwinia amy	242	6	1.2	205	1	UREG_KLEAE	P18319	klebsiella
170	6	1.2	146	1	HBE2_BOVIN	P06642	bos taurus	243	6	1.2	205	1	UREG_PROMI	O06306	proteus mir
171	6	1.2	146	1	HBE2_CAPIH	P02105	capra hircu	244	6	1.2	206	1	RUVA_RHETI	Q919c4	thizobium e
172	6	1.2	146	1	HBE4_BOVIN	P06643	bos taurus	245	6	1.2	206	1	UREG_SYNY3	P72955	synecocyst
173	6	1.2	147	1	RS12_METKA	O8cx12	methanopyru	246	6	1.2	206	1	UREG_UREPA	Q56361	ureaplasma
174	6	1.2	147	1	RS12_PYRAB	Q9v110	pyrococcus	247	6	1.2	208	1	INCB_ECOLI	P05631	ureaplasma
175	6	1.2	147	1	RS12_PYRHO	O59229	pyrococcus	248	6	1.2	209	1	LPCA_XENLA	P06692	xenopus lae
176	6	1.2	147	1	RS12_THECC	P29161	thermococci	249	6	1.2	209	1	LPCA_CHLTE	O8k4w3	chlorobium
177	6	1.2	148	1	CS52_THETH	P04164	thermus the	250	6	1.2	210	1	CUTI_PHYCP	P41754	phytophthor
178	6	1.2	148	1	CYC6_CHLRE	P08197	chlamydomon	251	6	1.2	210	1	VP28_CABEL	Q9n226	caenorhabdit
179	6	1.2	149	1	FUR_VIBAN	P37736	vibrio angu	252	6	1.2	211	1	END3_HAEIN	P44319	haemophilus

253	6	1.2	211	1	HI2_MOUSE	P15864	mus musculus	326	6	1.2	268	1	TRPA_PSEAE	P07344	pseudomonas
254	6	1.2	211	1	UREG_ACTPL	O54424	actinobacil	327	6	1.2	269	1	BASI_HUMAN	P35613	h basigin p
255	6	1.2	211	1	URK_BACDH	Q9Kd88	bacillus ha	328	6	1.2	269	1	TRPA_PSEPU	P11081	pseudomonas
256	6	1.2	212	1	ATPO_YEAST	P09457	saccharomyc	329	6	1.2	269	1	TRPA_PSESY	P34816	pseudomonas
257	6	1.2	212	1	YEAS_ECOLI	P76249	escherichia	330	6	1.2	269	1	TSRN_STRCN	P18644	streptomyces
258	6	1.2	213	1	HI3_FABIT	P02251	oryctolagus	331	6	1.2	270	1	Y041_CHLTR	O84045	chlamydia t
259	6	1.2	213	1	RB4A_HUMAN	P20338	homo sapien	332	6	1.2	270	1	Y382_RHIME	P25885	ribzobium m
260	6	1.2	213	1	RB4A_MOUSE	P50379	mus musculus	333	6	1.2	274	1	GADY_ECO57	P58820	escherichia
261	6	1.2	213	1	RB4A_RAT	P05714	rattus norv	334	6	1.2	274	1	GADY_ECOLI	P37639	escherichia
262	6	1.2	213	1	RB4B_HUMAN	P22750	homo sapien	335	6	1.2	274	1	NHS_STRAS	P52391	streptomyces
263	6	1.2	213	1	RB4B_RAT	P51146	rattus norv	336	6	1.2	275	1	PSBO_CYAAS	O96166	cyanothece
264	6	1.2	213	1	YLX_F_BACSU	P23454	bacillus su	337	6	1.2	276	1	RL2_BACHD	O92911	bacillus ha
265	6	1.2	214	1	PYRE_PASMU	Q9CJW4	pasteurella	338	6	1.2	277	1	KSGA_BURSP	O92911	burkholderi
266	6	1.2	216	1	IF6_THBAC	O06709	bordetella	339	6	1.2	279	1	TRPA_RHIME	O92911	rhizobium m
267	6	1.2	216	1	IF6_THBAC	O06709	bordetella	340	6	1.2	281	1	PANC_XYLA	O92911	xylothea fas
268	6	1.2	217	1	MMCI_METAC	P58867	methanosarc	341	6	1.2	283	1	GSPK_XANCP	P34056	xanthomonas
269	6	1.2	217	1	MMCI_METMA	P58977	methanosarc	342	6	1.2	283	1	SURE_CHLNU	O92911	chlamydia m
270	6	1.2	217	1	MMCI_METMA	P58977	methanosarc	343	6	1.2	283	1	SURE_CHLTR	O84220	chlamydia t
271	6	1.2	218	1	IF6_THBAC	P08284	gallus gall	344	6	1.2	284	1	HSF4_ARATH	O96320	arabidopsis
272	6	1.2	218	1	IF6_THBAC	P08284	gallus gall	345	6	1.2	284	1	YAGM_ECOLI	P71236	escherichia
273	6	1.2	218	1	IF6_THBAC	P58978	methanosarc	346	6	1.2	285	1	STAR_BOVIN	O28918	bos taurus
274	6	1.2	219	1	YDB2_SCHPO	O10355	schistosom	347	6	1.2	285	1	STAR_SHEEP	P79245	ovis aries
275	6	1.2	220	1	RNH2_BRIME	O68821	brucella me	348	6	1.2	287	1	PDX_ECOLI	P77150	escherichia
276	6	1.2	221	1	TRPF_HALVO	P52553	halobacteri	349	6	1.2	287	1	PDX_HAFLN	P44630	haemophilus
277	6	1.2	223	1	RB4A_ECHMU	O92911	echinococcu	350	6	1.2	289	1	HTPX_HAFLN	P44630	haemophilus
278	6	1.2	223	1	Y838_BACSU	O58248	bacillus su	351	6	1.2	289	1	PDX_PROMI	O51892	proteus mir
279	6	1.2	223	1	Y838_BACSU	O58248	methanococc	352	6	1.2	292	1	Y152_HUMAN	O14152	homo sapien
280	6	1.2	223	1	Y838_BACSU	P1851	caenorhabdi	353	6	1.2	294	1	EFTS_LISIN	O92911	listeria in
281	6	1.2	225	1	UREG_HAFLN	P44396	haemophilus	354	6	1.2	294	1	EFTS_LISMO	O84220	chlamydia m
282	6	1.2	226	1	NRF3_HAFLN	P45016	haemophilus	355	6	1.2	294	1	KC2B_CANAL	O92911	canalis
283	6	1.2	227	1	RNH2_AERPE	O92911	aeropyrum p	356	6	1.2	296	1	CEBB_MOUSE	P28033	mus musculus
284	6	1.2	230	1	PUR7_THEMA	O8XXX9	ralesonia s	357	6	1.2	296	1	CRK_XENLA	P87378	xenopus lae
285	6	1.2	230	1	TRPF_PALSO	O8XXX9	ralesonia s	358	6	1.2	296	1	NADC_ECOLI	P30011	escherichia
286	6	1.2	233	1	UBI4_MOUSE	P58321	mus musculu	359	6	1.2	296	1	NADC_RHORI	P77989	rhodospirill
287	6	1.2	233	1	YEB3_METUA	O58858	methanococc	360	6	1.2	296	1	NADC_SALTY	P30012	salmonella
288	6	1.2	235	1	EFTU_COSCS	P50376	costaria co	361	6	1.2	296	1	URER_ECOLI	P32336	escherichia
289	6	1.2	235	1	EFTU_GLOSI	P50063	gloeotheca	362	6	1.2	297	1	HSLO_PSEAE	O92911	pseudomonas
290	6	1.2	235	1	EFTU_PLEBO	O56332	pleococcusa	363	6	1.2	298	1	HMGL_CHICK	P35915	gallus gall
291	6	1.2	237	1	NAPC_PARNB	O9X017	thermotoga	364	6	1.2	298	1	Y535_UREPA	O92911	ureaplasma
292	6	1.2	238	1	RNH2_THEMA	P22405	streptomyces	365	6	1.2	300	1	MTG9_CITFR	P40388	bacillus su
293	6	1.2	241	1	KORA_STRLI	O92911	serinus can	366	6	1.2	300	1	MTX1_XANCC	P14243	citrobacter
294	6	1.2	241	1	NEUM_SERCA	P37319	herpes simp	367	6	1.2	300	1	SYGA_PASMU	P50794	pasteurella
295	6	1.2	245	1	ICP3_HSV1N	P42216	escherichia	368	6	1.2	301	1	ERA_ECO57	P58070	escherichia
296	6	1.2	245	1	KSUS_ECOLI	P38773	saccharomyc	370	6	1.2	301	1	ERA_ECOLI	P06616	escherichia
297	6	1.2	246	1	DOG2_YEAST	P38773	saccharomyc	371	6	1.2	301	1	ERA_SALTY	O82457	salmonella
298	6	1.2	246	1	PSAS_TRIBB	O9X255	cypanosoma	372	6	1.2	301	1	ERA_SALTY	O82457	salmonella
299	6	1.2	247	1	FAG1_SYNY3	P73574	synechocyst	373	6	1.2	301	1	SYGA_NEIMA	O92911	neisseria m
300	6	1.2	247	1	TFAD_ECOLI	P77659	escherichia	374	6	1.2	301	1	VNSM_TSMVI	P31654	pancoea ana
301	6	1.2	248	1	ICP3_HSV11	P36313	herpes simp	375	6	1.2	302	1	CRTE_PANAN	P43758	haemophilus
302	6	1.2	250	1	URBF_UREPA	O56560	ureaplasma	377	6	1.2	302	1	PTB_CLOBE	O05624	clostridium
303	6	1.2	250	1	A32B_HUMAN	O92911	homo sapien	378	6	1.2	302	1	SYGA_HAFLN	P43821	haemophilus
304	6	1.2	251	1	RS6_ASPOF	O92911	asparagus o	379	6	1.2	303	1	CRKL_HUMAN	P46109	homo sapien
305	6	1.2	251	1	TPIS_PSEAE	O92911	pseudomonas	380	6	1.2	303	1	CRKL_MOUSE	P47941	mus musculus
306	6	1.2	251	1	TPIS_PSEAE	O92911	pseudomonas	381	6	1.2	303	1	ERA_YERPE	O82457	yerkesia pe
307	6	1.2	252	1	ICP3_HSV1D	P37318	herpes simp	382	6	1.2	303	1	FMT_RICCN	O33519	ricketsia
308	6	1.2	252	1	ICP3_HSV1D	P37318	herpes simp	383	6	1.2	303	1	VNSM_TSMVI	O01268	impatiens n
309	6	1.2	252	1	ICP3_HSV1D	P37318	herpes simp	384	6	1.2	304	1	CRK_HUMAN	P46109	homo sapien
310	6	1.2	253	1	CB2_DUNTE	P27517	dunalifella	385	6	1.2	304	1	CRK_MOUSE	O64010	mus musculus
311	6	1.2	253	1	NCAP_TOSV	P21701	toscana vir	386	6	1.2	304	1	CRK_MOUSE	O63768	rattus norv
312	6	1.2	253	1	TATC_AZOCN	P40855	azotobacter	387	6	1.2	305	1	Y007_MYCTU	P15155	mycobacteri
313	6	1.2	255	1	YPMR_BACSU	P40766	bacillus su	388	6	1.2	305	1	YEHF_ECOLI	P33362	escherichia
314	6	1.2	255	1	YPMR_BACSU	P40766	bacillus su	389	6	1.2	306	1	ERA_PASMU	O92911	pasteurella
315	6	1.2	258	1	RNH2_LACIA	O92911	laetococcus	390	6	1.2	306	1	HUTG_HALNI	O92911	halobacteri
316	6	1.2	258	1	SURE_SCAMU	O92911	campylobact	391	6	1.2	308	1	FMT_NEIMA	O92911	neisseria m
317	6	1.2	258	1	YDF2_SCHPO	O10474	schistosom	392	6	1.2	309	1	PDXK_ARATH	O84220	arabidopsis
318	6	1.2	259	1	DEFM_ARATH	O92911	arabidopsis	393	6	1.2	309	1	MCAL_MOUSE	P31230	mus musculus
319	6	1.2	259	1	TRUA_METUA	O92911	methanococc	394	6	1.2	310	1	ODCI_YEAST	O03028	saccharomyc
320	6	1.2	261	1	ICP3_HSV1F	P08353	herpes simp	395	6	1.2	310	1	BLAC_STRAU	P10509	streptomyces
321	6	1.2	264	1	LDLR_BOVIN	P01131	bos taurus	396	6	1.2	311	1	MSAB_STRGC	O91269	s peptide m
322	6	1.2	267	1	AXII_ARATH	O24408	arabidopsis	397	6	1.2	311	1	PDXK_HUMAN	O00764	homo sapien
323	6	1.2	267	1	AXII_ARATH	O24408	arabidopsis	398	6	1.2	312	1	PDXK_HUMAN	O00764	homo sapien

399	6	1.2	312	1	PDXK_SHEEP	P82197	ovis aries	472	6	1.2	359	1	YKHS_CABEL	P34273	caenorhabdi
400	6	1.2	312	1	Y426_ARCFU	C29823	archaeoglob	473	6	1.2	360	1	MOA2_MYCTU	053881	mycobacteri
401	6	1.2	312	1	YEC9_YEAST	P39988	saccharomyc	474	6	1.2	360	1	Y303_HELPU	092083	helicobacte
402	6	1.2	315	1	YNEF_ECO57	Q8X242	escherichia	475	6	1.2	360	1	Y303_HELPU	025074	helicobacte
403	6	1.2	315	1	YNEF_ECOLI	P76147	escherichia	476	6	1.2	361	1	COBT_MYCTU	010306	mycobacteri
404	6	1.2	316	1	YH58_AQUAE	O67638	aquilex aeo	477	6	1.2	361	1	LIG_PHLA	P20010	phlebia rad
405	6	1.2	316	1	YK95_MYCTU	Q10704	mycobacteri	478	6	1.2	362	1	YBDB_ECOLI	P45579	escherichia
406	6	1.2	317	1	YK95_YEAST	P53727	saccharomyc	479	6	1.2	363	1	ALFB_HUMAN	P05082	homo sapien
407	6	1.2	319	1	TRXB_BUCAP	P57399	buchnera ap	480	6	1.2	364	1	VAOD_NEUCR	P53659	neutrospora
408	6	1.2	319	1	TRXB_BUCAP	P81433	buchnera ap	481	6	1.2	364	1	YM31_MYCTU	010503	mycobacteri
409	6	1.2	321	1	PDXK_CABEL	O01824	caenorhabdi	482	6	1.2	365	1	FLC1_PROMI	P42272	proteus mir
410	6	1.2	324	1	GER_THENT	Q81b77	thermoanaer	483	6	1.2	365	1	RF2_LACLA	Q9G9X1	lactococcus
411	6	1.2	324	1	YK95_MYCLE	P54073	mycobacteri	484	6	1.2	365	1	SUCC_METTH	027115	methanobact
412	6	1.2	325	1	HMGL_RAT	P97519	rattus norv	485	6	1.2	365	1	YM27_YEAST	P28652	saccharomyc
413	6	1.2	325	1	SECF_HAEIN	P45590	haemophilus	486	6	1.2	366	1	IC13_HUMAN	P30506	homo sapien
414	6	1.2	327	1	NIFH_CYAA5	O07641	cyanothece	487	6	1.2	366	1	VUI_HAV6U	001349	human herpe
415	6	1.2	329	1	COA4_STRCO	O86779	streptomyce	488	6	1.2	366	1	YVAF_BACSU	P37518	bacillus su
416	6	1.2	329	1	Y493_MYCTU	Q11158	mycobacteri	489	6	1.2	368	1	SSR2_BOVIN	P34993	bos taurus
417	6	1.2	331	1	PSO2_ARATH	Q98841	arabidopsis	490	6	1.2	368	1	Y812_AQUAE	O66994	aquilex aeo
418	6	1.2	331	1	SP2B_BACSU	P37575	bacillus su	491	6	1.2	369	1	FORL_BUCAP	O8K918	buchnera ap
419	6	1.2	333	1	ILVC_MYCAV	O59500	mycobacteri	492	6	1.2	370	1	DNAJ_BACHD	O9Kd71	bacillus ha
420	6	1.2	333	1	ILVC_MYCLE	O33114	mycobacteri	493	6	1.2	370	1	RECF_BACSU	P05651	bacillus su
421	6	1.2	333	1	ILVC_MYCTU	O53248	mycobacteri	494	6	1.2	371	1	ISDF_CAMJE	O9Pm68	c isepd/ispf
422	6	1.2	334	1	FRUR_ECOLI	P21168	escherichia	495	6	1.2	372	1	FLGI_XANAC	O8P127	xanthomonas
423	6	1.2	334	1	FRUR_SALTY	P21930	salmonella	496	6	1.2	372	1	OPRD_XANCP	O8P960	xanthomonas
424	6	1.2	334	1	RVUB_BACSU	O32055	bacillus su	497	6	1.2	372	1	OPRD_HUMAN	P41143	homo sapien
425	6	1.2	336	1	RVUB_METTH	O27717	methanobact	498	6	1.2	372	1	OPRD_MOUSE	P32300	mus musculus
426	6	1.2	337	1	COBU_PSDDE	P29935	pseudomonas	499	6	1.2	372	1	OPRD_RAT	P33533	rattus norv
427	6	1.2	337	1	GPI1_GIALA	P53429	giardia lam	500	6	1.2	375	1	GCST_AARPE	O9Yb22	aeropyrum p
428	6	1.2	337	1	ILVC_COREF	Q81P41	corynebacte	501	6	1.2	376	1	G4P1_YEAST	P46672	saccharomyc
429	6	1.2	337	1	YF02_CLOPE	P58698	clostridium	502	6	1.2	376	1	NIRK_RHIME	Q92249	rhizobium m
430	6	1.2	338	1	DCUP_AQUAE	O66677	aquilex aeo	503	6	1.2	377	1	DNAJ_BRALJA	P94319	bradyrhizob
431	6	1.2	338	1	ILVC_CORGL	O57179	corynebacte	504	6	1.2	377	1	GP27_RAT	O91j33	rattus norv
432	6	1.2	340	1	DCUP_CAMJE	O9Pm54	campylobact	505	6	1.2	377	1	RNGI_HUMAN	O06587	homo sapien
433	6	1.2	340	1	MYBI_MAIZE	P20024	zea mays (m	506	6	1.2	378	1	ELVA_BACSP	P20724	bacillus sp
434	6	1.2	340	1	RECA_THETH	P48297	thermus the	507	6	1.2	378	1	MOBL_THIPE	P20085	thiobacilli
435	6	1.2	341	1	CYTR_ECOLI	P09664	escherichia	508	6	1.2	379	1	GP27_MOUSE	O54809	mus musculus
436	6	1.2	342	1	COBT_METSO	Q9X7E4	methylobact	509	6	1.2	379	1	NIR_RHIGA	O01537	rhizobium g
437	6	1.2	343	1	DHSO_BACHD	Q929U1	bacillus ha	510	6	1.2	380	1	DNAJ_BACST	O45552	bacillus st
438	6	1.2	343	1	HMD_METVO	O50840	methanococc	511	6	1.2	382	1	GRK_BACSU	P42100	bacillus su
439	6	1.2	343	1	RADA_HALVO	Q48328	halodacteri	512	6	1.2	382	1	KCCI_METAN	O14408	metathium
440	6	1.2	344	1	HMD_METTF	Q50758	m coenzyme	513	6	1.2	385	1	SNFS_HUMAN	Q12824	homo sapien
441	6	1.2	344	1	HMD_METTH	O27211	m coenzyme	514	6	1.2	385	1	SNFS_MOUSE	Q92083	mus musculus
442	6	1.2	344	1	HMD_METTH	P32440	m coenzyme	515	6	1.2	386	1	YBDL_ECOLI	P77806	escherichia
443	6	1.2	344	1	HMD_METTH	P81221	m coenzyme	516	6	1.2	388	1	NORA_STRAAM	P21311	staphylococ
444	6	1.2	344	1	KAND_SCHPO	O10078	escherichia	517	6	1.2	388	1	SUCC_COXBU	P53582	coxiella bu
445	6	1.2	345	1	HRCA_LISIN	Q922B6	listeria in	518	6	1.2	388	1	G3PA_TOBAC	P09043	nicotiana t
446	6	1.2	345	1	HRCA_LISMO	Q95566	listeria mo	519	6	1.2	388	1	YVAT_YEAST	P31412	saccharomyc
447	6	1.2	346	1	FLIG_AGR75	P45484	rhizobium m	520	6	1.2	392	1	CYB_METSE	O47489	metridium s
448	6	1.2	347	1	PUR7_HAINT	Q44458	agrobacteri	521	6	1.2	393	1	YEGO_YEAST	P32611	saccharomyc
449	6	1.2	347	1	Y576_METJA	Q9Hn77	halobacteri	522	6	1.2	393	1	FADH_PSEPU	P46154	pseudomonas
450	6	1.2	347	1	Y576_METJA	O57996	methanococc	523	6	1.2	398	1	HRCA_CHLPP	O92860	chlamydia p
451	6	1.2	347	1	Y576_METJA	P34583	caenorhabdi	524	6	1.2	398	1	MPK2_CHICK	O90801	gallus gall
452	6	1.2	348	1	AROH_ERWHE	O54459	erwinia her	525	6	1.2	398	1	NIS2_ANAVA	O44482	anabena va
453	6	1.2	348	1	RTCA_ANASP	O8Yv83	anabena sp	526	6	1.2	399	1	EFTU_CAMJE	O69303	campylobact
454	6	1.2	350	1	Y098_SYNY3	O65880	synechocyst	527	6	1.2	400	1	CN37_BOVIN	P06633	bos taurus
455	6	1.2	351	1	NOEL_RHIFR	O85713	rhizobium f	528	6	1.2	400	1	MPK2_HUMAN	P36507	homo sapien
456	6	1.2	351	1	NOEL_RHIFR	P55354	rhizobium s	529	6	1.2	400	1	MPK2_RAT	P36506	rattus norv
457	6	1.2	353	1	GAG_MMTVG	P03343	mouse mamma	530	6	1.2	400	1	PGK_CAMJE	O9Pm65	campylobact
458	6	1.2	353	1	HMD_METTL	Q50759	m coenzyme	531	6	1.2	400	1	PUR2_THBMA	O9X0X7	thermotoga
459	6	1.2	354	1	E2B_XYLPA	Q9Paf5	xyliella fas	532	6	1.2	401	1	MPK2_MOUSE	O63932	mus musculus
460	6	1.2	354	1	MURG_ECO57	O8X9V8	escherichia	533	6	1.2	402	1	ARGJ_METJA	O57665	methanococc
461	6	1.2	354	1	MURG_ECOLI	O8E164	escherichia	534	6	1.2	402	1	CSD_UREPA	O9Pq36	ureaplaasma
462	6	1.2	354	1	MURG_ECOLI	P17443	escherichia	535	6	1.2	402	1	FIXF_RHISN	P55467	rhizobium s
463	6	1.2	354	1	YGPI_YEAST	P38616	saccharomyc	536	6	1.2	403	1	ARGG_CAVCR	O9A3J4	c arginine
464	6	1.2	355	1	GBAC_CABEL	Q19572	caenorhabdi	537	6	1.2	404	1	HMDH_HALVO	O59498	halobacteri
465	6	1.2	355	1	MURG_NEIMA	O9J577	neisseria m	538	6	1.2	404	1	CGE2_HUMAN	O92020	mus musculus
466	6	1.2	355	1	MURG_NEIMA	O9K0Z2	neisseria m	539	6	1.2	404	1	CGE2_MOUSE	O92027	archaeoglob
467	6	1.2	356	1	YD6B_ASFB7	P03325	african swi	540	6	1.2	404	1	HISX_ARCFU	O57484	haemophilus
468	6	1.2	356	1	YD6B_SCHPO	Q10323	schizosacch	541	6	1.2	404	1	Y867_HAEIN	O9Pdk4	xyliella fas
469	6	1.2	357	1	KARG_ERISI	Q9Hn48	erichocheit s	542	6	1.2	405	1	TRB8_XYLPA	P38746	saccharomyc
470	6	1.2	358	1	HMD_METJA	O58194	m coenzyme	543	6	1.2	405	1	YLP2_YEAST		
471	6	1.2	358	1	PONI_RABIT	P27170	oryzocolagus	544	6	1.2	408	1	APGI_ARCFU	O28523	archaeoglob

545	6	1.2	409	1	EFTU_ASTLO	P14634	astasia lon	618	6	1.2	463	1	ZABH_SCHPO	Q12702	schizosacch
546	6	1.2	409	1	EFTU_CVAPA	P17245	cyamophora	619	6	1.2	463	1	YIBO_HAEIN	P44903	haemophilus
547	6	1.2	409	1	EFTU_EUGGR	P02991	euglena gra	620	6	1.2	464	1	FLID_AERPU	Q91976	aeromonas p
548	6	1.2	410	1	EFTU_CVACA	O9c138	cyanidium c	621	6	1.2	464	1	YTI3_CAEBL	Q10917	caenorhabdi
549	6	1.2	412	1	CLPX_AOUAE	O67356	agufex aeo	622	6	1.2	465	1	SNX8_HUMAN	Q9Y552	homo sapien
550	6	1.2	412	1	DAD3_RHILQ	O98182	rhizobium 1	623	6	1.2	467	1	MATR_RAT	P41243	rattus norv
551	6	1.2	412	1	DJAD2_HUMAN	O60864	homo sapien	624	6	1.2	467	1	NX1B_BOVIN	Q28172	bos taurus
552	6	1.2	412	1	DJAD2_MOUSE	O9qy10	mus musculu	625	6	1.2	467	1	SYC_THETN	Q84763	thermomae
553	6	1.2	412	1	DJAD2_RAT	O35824	rattus norv	626	6	1.2	468	1	6PGB_BACSU	P12013	bacillus su
554	6	1.2	413	1	PBP2_YEAST	P38181	saccharomyc	627	6	1.2	468	1	NX1B_RAT	O63337	rattus norv
555	6	1.2	413	1	PROA_CAICR	O9a2x6	caulobacter	628	6	1.2	469	1	CHLN_SYNY3	P28172	synchoecyt
556	6	1.2	414	1	CPXA_PSEPU	P00183	pseudomonas	629	6	1.2	469	1	TRPC_PASMU	P57855	pasteurella
557	6	1.2	414	1	KCCI_EMENT	O00771	emeritella	630	6	1.2	470	1	NR54_HUMAN	Q15233	homo sapien
558	6	1.2	415	1	PAX2_MOUSE	P32114	mus musculu	631	6	1.2	471	1	MURD_STRCO	O62499	streptomyce
559	6	1.2	416	1	PAX2_HUMAN	O02962	homo sapien	632	6	1.2	472	1	KPYK_MYCTU	O06134	mycobacteri
560	6	1.2	416	1	RAGE_BOVIN	O28173	bos taurus	633	6	1.2	472	1	PLS8_SPIOL	Q34657	spiniace o1
561	6	1.2	417	1	GLYA_BORBU	O51547	borrelia bu	634	6	1.2	474	1	NUAM_PARTE	P15581	paramecium
562	6	1.2	417	1	BGK_CANMA	P41757	candida mal	635	6	1.2	476	1	EFTU_ARATH	P17745	arabidopsis
563	6	1.2	418	1	PROA_CIOAB	O97682	clostridium	636	6	1.2	476	1	TH11_THEVO	Q97466	thermoplas
564	6	1.2	419	1	CLPX_CHLTR	O84711	chlamydia t	637	6	1.2	477	1	MURC_XANAC	O8p347	xanthomonas
565	6	1.2	420	1	CN37_MOUSE	P16330	mus musculu	638	6	1.2	477	1	MURC_XANCP	O8pc39	xanthomonas
566	6	1.2	421	1	CN37_HUMAN	P09543	homo sapien	639	6	1.2	477	1	MURC_XYLFA	O9f850	xyella fas
567	6	1.2	422	1	ADHF_SCHPO	O09659	schizosacch	640	6	1.2	478	1	HRTS_CROAT	P34182	crotaeus at
568	6	1.2	424	1	CLPX_RALSO	O8xy06	raistonia s	641	6	1.2	478	1	MAOX_BACST	P16468	bacillus st
569	6	1.2	424	1	PUR2_HELPU	O9zk07	helicobacte	642	6	1.2	478	1	T23C_BACTB	O8p347	xanthomonas
570	6	1.2	424	1	PUR2_HELPY	O25817	helicobacte	643	6	1.2	479	1	EFTI_SOYBN	Q43467	glycine max
571	6	1.2	424	1	VID2_AGRTR	P06668	agrobacteri	644	6	1.2	479	1	P2CB_HUMAN	O75688	homo sapien
572	6	1.2	425	1	SYS_STRCO	O9zbx1	streptomyce	645	6	1.2	480	1	NR61_HUMAN	O15405	homo sapien
573	6	1.2	426	1	PROA_RALSO	O8xy66	raistonia s	646	6	1.2	482	1	BHB3_HUMAN	O9c039	homo sapien
574	6	1.2	426	1	ST25_HUMAN	O00506	homo sapien	647	6	1.2	483	1	MURC_RALSO	O8xy18	raistonia s
575	6	1.2	426	1	ST25_MOUSE	O9z2w1	mus musculu	648	6	1.2	483	1	TRKH_ECOLI	P21166	escherichia
576	6	1.2	427	1	HISX_STRMU	O8dtq7	streptococc	649	6	1.2	483	1	TRKH_ECOLI	Q91616	escherichia
577	6	1.2	427	1	TIG_LACLA	O9c115	lactococcu	650	6	1.2	484	1	TH11_PSEAE	O9h569	pseudomonas
578	6	1.2	427	1	TIG_STRPN	O97869	streptococc	651	6	1.2	484	1	UDPE_NPVOP	O65163	oryzia pseu
579	6	1.2	428	1	ENPE_HUMAN	O75356	homo sapien	652	6	1.2	491	1	YOJ6_CAEBL	P34629	caenorhabdi
580	6	1.2	428	1	PROA_STRCO	O97dK1	scieptomyce	653	6	1.2	492	1	FSCI_HUMAN	Q16658	homo sapien
581	6	1.2	429	1	COBB_RHIME	O92p48	rhizobium m	654	6	1.2	492	1	FSCI_MOUSE	O61658	mus musculu
582	6	1.2	429	1	GSAB_BACSU	P71084	bacillus su	655	6	1.2	493	1	DCMC_HUMAN	O95882	homo sapien
583	6	1.2	429	1	LAMB_KLEPN	P31242	klebsiella	656	6	1.2	494	1	NORV_VIBVU	O8d4f8	vibrio vuln
584	6	1.2	430	1	NY42_YEAST	P49686	saccharomyc	657	6	1.2	495	1	NR61_MOUSE	O64239	mus musculu
585	6	1.2	430	1	SYMC_YEAST	Q12109	saccharomyc	658	6	1.2	495	1	TRME_TREPA	O83561	treponema p
586	6	1.2	431	1	COBB_PSEDE	P21632	pseudomonas	659	6	1.2	497	1	AMPA_RHIME	O92qy7	thizobium p
587	6	1.2	434	1	CISY_BRAJA	P94335	bradyrhizob	660	6	1.2	498	1	DG11_RAT	O9emf3	rattus norv
588	6	1.2	435	1	PURA_CHLTE	O8kak6	chloobium	661	6	1.2	498	1	PR1M_AOUAE	O67465	agufex aeo
589	6	1.2	436	1	Y868_CHLMU	O9p3j1	chlamydia m	662	6	1.2	498	1	VGLX_PRIVT	P07562	pseudorabie
590	6	1.2	437	1	CTBO_XENLA	O9w758	xenopus lae	663	6	1.2	499	1	CIW5_HUMAN	O95279	homo sapien
591	6	1.2	439	1	AROA_DEIRA	O9rvd3	deinococcus	664	6	1.2	499	1	CPN1_RAT	P13393	rattus norv
592	6	1.2	440	1	BCHN_HEIMO	O92969	heliobacill	665	6	1.2	500	1	IC1_HUMAN	P05155	homo sapien
593	6	1.2	440	1	GAGC_AYISC	P05433	avian sarco	666	6	1.2	501	1	PYCA_METJA	O58626	methanococc
594	6	1.2	440	1	YHJE_ECOLI	P37643	escherichia	667	6	1.2	502	1	SPHM_HUMAN	P51688	homo sapien
595	6	1.2	442	1	NX1B_HUMAN	P58400	homo sapien	668	6	1.2	503	1	AMVH_WHEAT	P93594	tritlicum ae
596	6	1.2	443	1	MT17_YEAST	P06106	saccharomyc	669	6	1.2	503	1	FUT4_DROME	O9vui9	drosophila
597	6	1.2	443	1	MURC_LACLA	O9ce10	lactococcu	670	6	1.2	503	1	NRP1_HUMAN	Q16656	homo sapien
598	6	1.2	446	1	KCCI_YEAST	P27466	saccharomyc	671	6	1.2	503	1	NRP1_MOUSE	O9w400	mus musculu
599	6	1.2	447	1	KCC2_YEAST	P22517	saccharomyc	672	6	1.2	504	1	FLIC_SALBU	O06596	salmonella
600	6	1.2	448	1	PO31_HUMAN	O03052	homo sapien	673	6	1.2	504	1	FLIC_SALDE	O06970	salmonella
601	6	1.2	449	1	LAT_MYCTU	P96895	mycobacteri	674	6	1.2	504	1	FLIC_SALDU	O06971	salmonella
602	6	1.2	449	1	PO31_MOUSE	P21952	mus musculu	675	6	1.2	504	1	FLIC_SALEN	O06972	salmonella
603	6	1.2	450	1	LAT_NOCIA	O05174	nocardia la	676	6	1.2	504	1	FLIC_SALWC	O06981	salmonella
604	6	1.2	451	1	PO31_RAT	P20267	rattus norv	677	6	1.2	504	1	FLIC_SALWA	O06973	salmonella
605	6	1.2	451	1	YOEV_BACSU	P54462	bacillus su	678	6	1.2	504	1	FLIC_SALNA	O52359	salmonella
606	6	1.2	452	1	YP59_MYCTU	O50739	mycobacteri	679	6	1.2	504	1	FLIC_SALRO	O06982	salmonella
607	6	1.2	454	1	RHBE_RHIME	O9z3g8	rhizobium m	680	6	1.2	504	1	FLIC_SALSE	O06983	salmonella
608	6	1.2	454	1	UCR2_NEUCR	O60044	neutrospora	681	6	1.2	504	1	IC1_MOUSE	P97290	mus musculu
609	6	1.2	456	1	MANB_SALMO	O01411	salmonella	682	6	1.2	504	1	NRP1_BEAR	O90544	brachydanio
610	6	1.2	456	1	MANB_SALTY	P26341	bos taurus	683	6	1.2	505	1	MARK_MOUSE	P41442	mus musculu
611	6	1.2	456	1	NX3B_BOVIN	O28143	bos taurus	684	6	1.2	505	1	MURE_SYNY3	O55469	synchoecyt
612	6	1.2	457	1	BPHA_COMTE	O46372	comamonas t	685	6	1.2	506	1	PALA_PPAVU	P07218	phaseolus v
613	6	1.2	457	1	LAT_STRCL	O01767	streptomyce	686	6	1.2	507	1	C7C3_ARATH	O64658	arabidopsis
614	6	1.2	459	1	Y976_TREPA	O83941	treponema p	687	6	1.2	507	1	FLIC_SALBE	O06974	salmonella
615	6	1.2	460	1	EFIA_NEUCR	O01372	neutrospora	688	6	1.2	507	1	FLIC_SALON	O06974	salmonella
616	6	1.2	460	1	HEMO_MOUSE	O91x72	mus musculu	689	6	1.2	507	1	MATR_HUMAN	P42679	homo sapien
617	6	1.2	460	1	HEMO_RABIT	P20058	oryctolagus	690	6	1.2	508	1	PSBB_EUGGR	P14813	euglena gra

691	1.2	508	1	RPA2	THEAC	Q03586	thermoplasm	764	6	1.2	576	1	MOES	HUMAN	P26038	homo sapien
692	1.2	509	1	GUAA	CLOPE	Q81466	clostridium	765	6	1.2	576	1	MOES	MOUSE	P26041	mus musculus
693	1.2	510	1	BMP6	MOUSE	P20722	mus musculus	766	6	1.2	576	1	MOES	PIG	P26042	mus scrofa
694	1.2	510	1	GPMT	PSESMS	P52832	pseudomonas	767	6	1.2	576	1	SBP	CAREL	Q21950	caenorhabditis
695	1.2	511	1	AROF	LYCES	P37215	lycopericon	768	6	1.2	577	1	SYR	VIRCH	Q9K966	vibrio chol
696	1.2	511	1	AROG	SOLTU	P37822	solianum tub	769	6	1.2	578	1	DSBD	PSEBP	Q9K123	pseudomonas
697	1.2	511	1	COPD	BOVIN	P53619	bos taurus	770	6	1.2	579	1	SYQ	XANAC	Q8pn35	xanthomonas
698	1.2	511	1	COPD	HUMAN	P48444	homo sapien	771	6	1.2	579	1	SYQ	XANAC	Q8pn35	xanthomonas
699	1.2	512	1	C26A	HUMAN	Q9n636	homo sapien	772	6	1.2	580	1	DAKI	SCHPO	Q13992	schizosacch
700	1.2	513	1	HUTH	RHILLO	Q93813	rhizobium l	773	6	1.2	580	1	EZRI	BOVIN	P31976	bos taurus
701	1.2	514	1	VG64	HSV1	Q00156	ictaluriid h	774	6	1.2	580	1	MEND	BACSU	P23970	b menaquinone
702	1.2	516	1	BOCT	RAT	Q9P290	rattus norv	775	6	1.2	583	1	KPIA	RICCO	Q43117	ricinus com
703	1.2	516	1	CE23	HORVU	P52711	hordeum vul	776	6	1.2	583	1	RADI	HUMAN	P35241	homo sapien
704	1.2	522	1	MSAB	NEIGO	P14930	n peptidic m	777	6	1.2	583	1	RADI	MOUSE	P26043	mus musculus
705	1.2	522	1	MSAB	NEIMA	Q9jmb6	n peptidic m	778	6	1.2	583	1	RADI	PIG	P26044	mus scrofa
706	1.2	522	1	MSAB	NEIMB	Q9klm8	n peptidic m	779	6	1.2	584	1	YMB3	YEAST	Q04228	saccharomyc
707	1.2	524	1	GUAA	YEAST	P38625	saccharomyc	780	6	1.2	585	1	EZRI	HUMAN	P15311	homo sapien
708	1.2	527	1	TP6B	PYRAE	Q8zvm0	pyrodaculum	781	6	1.2	585	1	EZRI	MOUSE	P26040	mus musculus
709	1.2	528	1	FOLE	NEOCR	O13492	neurospora	782	6	1.2	586	1	SYR	STRCO	Q9wx29	strepitomyce
710	1.2	529	1	NMT	AJECA	P34763	atellomyces	783	6	1.2	587	1	DSD2	PSEAB	Q91104	pseudomonas
711	1.2	529	1	NX1A	MOUSE	Q9c684	mus musculus	784	6	1.2	587	1	YNI4	YEAST	P53933	saccharomyc
712	1.2	529	1	TIWK	ECOLI	P08957	escherichia	785	6	1.2	588	1	ASH1	YEAST	P34223	saccharomyc
713	1.2	529	1	TIW	SALPO	P07989	salmonella	786	6	1.2	589	1	STER	CANAL	P46599	candida alb
714	1.2	529	1	TIW	SALTY	P40813	salmonella	787	6	1.2	591	1	GAG	MTVB	P10258	mouse mamma
715	1.2	530	1	UD12	HUMAN	P36509	homo sapien	788	6	1.2	593	1	KPIA	TOBAC	Q40545	nicotiana t
716	1.2	530	1	UD17	HUMAN	Q9haw7	homo sapien	789	6	1.2	594	1	DCP2	SCHPO	Q92345	schewanelia
717	1.2	530	1	UD18	HUMAN	Q9haw7	homo sapien	790	6	1.2	596	1	FRDA	SHEON	P83223	schewanelia
718	1.2	530	1	UD19	HUMAN	O60656	homo sapien	791	6	1.2	596	1	TFB8	YEAST	P29056	saccharomyc
719	1.2	530	1	UD1A	HUMAN	Q9haw7	homo sapien	792	6	1.2	599	1	TYPA	HELPU	Q92123	helicobacte
720	1.2	531	1	KEYK	ELITE	O44006	elmeria ten	793	6	1.2	599	1	TYPA	HELPU	O25225	helicobacte
721	1.2	532	1	UV16	HUMAN	P19224	homo sapien	794	6	1.2	600	1	DCP1	ORYSA	Q92755	chlamydia p
722	1.2	533	1	UV1N	RHILR	O05467	rhizobium t	795	6	1.2	602	1	DCP1	ORYSA	P51847	oryza sativ
723	1.2	533	1	UD11	HUMAN	P22310	homo sapien	796	6	1.2	603	1	RPB1	METTH	O27124	methanobact
724	1.2	534	1	K2C4	HUMAN	P19013	homo sapien	797	6	1.2	604	1	RPB1	METTH	P09845	methanobact
725	1.2	534	1	UD13	HUMAN	P35503	homo sapien	798	6	1.2	605	1	NUEM	PELSU	O79678	pelomedusa
726	1.2	534	1	UD14	HUMAN	P22310	homo sapien	799	6	1.2	607	1	GDST	HUMAN	P52306	homo sapien
727	1.2	534	1	UD15	HUMAN	P35503	homo sapien	800	6	1.2	609	1	PPCK	MYCLE	O06084	mycobacteri
728	1.2	536	1	CH62	CHLVC	P59698	chlamydomo	801	6	1.2	610	1	ARLC	MAIZE	P13526	zea mays (m
729	1.2	538	1	LY41	THEVA	O94222	thermotoga	802	6	1.2	611	1	ACFS	PELCA	O62763	fella silve
730	1.2	539	1	CH60	FUSUN	O8r575	fusobacteri	803	6	1.2	612	1	ARBS	MAIZE	P13077	zea mays (m
731	1.2	539	1	CH60	FUSUN	O8r575	fusobacteri	804	6	1.2	613	1	ACFS	BOVIN	P23795	bos taurus
732	1.2	541	1	GT10	HUMAN	O8g100	fusobacteri	805	6	1.2	614	1	ILVD	RHILLO	Q98028	rhizobium l
733	1.2	543	1	CH60	BRECH	O95528	homo sapien	806	6	1.2	619	1	ABS	DROME	Q9v3c0	drosophila
734	1.2	544	1	CH63	RHILE	Q8rui0	brevibacilli	807	6	1.2	622	1	PPID	HAELN	P44092	haemophilus
735	1.2	546	1	AAAS	HUMAN	O91690	rhizobium l	808	6	1.2	623	1	Y711	HUMAN	P00550	escherichia
736	1.2	546	1	AAAS	HUMAN	O9n979	homo sapien	809	6	1.2	625	1	NIFA	AZOBR	P10912	homo sapien
737	1.2	548	1	WR72	ARATH	P58742	mus musculus	810	6	1.2	628	1	LU	HUMAN	P50895	homo sapien
738	1.2	552	1	ILVD	CLOAB	Q91x98	arabidopsis	811	6	1.2	629	1	PAB2	ARATH	P42731	arabidopsis
739	1.2	554	1	GA44	HUMAN	O97ee3	clostridium	812	6	1.2	632	1	PABP	DROME	P21877	drosophila
740	1.2	556	1	GCH2	SYNY3	P48169	homo sapien	813	6	1.2	635	1	PTMA	KLEPN	Q9xbm7	klesiella
741	1.2	556	1	PT1	SMRGO	P74104	synechocyst	814	6	1.2	637	1	PTMA	ECOLI	P00550	escherichia
742	1.2	556	1	SVQ	PSEAE	O9k2p1	streptomyce	815	6	1.2	638	1	GHR	HUMAN	P10912	homo sapien
743	1.2	557	1	R81	ECOLI	Q912u8	pseudomonas	816	6	1.2	638	1	GHR	MACMU	P79194	macaca mula
744	1.2	558	1	GDST	BOVIN	P02349	escherichia	817	6	1.2	643	1	CTR2	XENLA	P79955	xenopus lae
745	1.2	559	1	THSA	SULSH	Q04173	bos taurus	818	6	1.2	643	1	K2C1	HUMAN	P04264	homo sapien
746	1.2	560	1	THSA	SULSH	O9v269	sulfolobus	819	6	1.2	643	1	SENI	HUMAN	Q9p0u3	homo sapien
747	1.2	562	1	CHS5	CANAL	P46129	sulfolobus	820	6	1.2	643	1	Y903	METVA	O58313	methanococ
748	1.2	562	1	DHE3	DROME	O74161	candida alb	821	6	1.2	644	1	HS71	PICAN	P53421	metaphano
749	1.2	563	1	Y288	CHLTR	P54385	drosophila	822	6	1.2	644	1	PLB1	SCHPO	P53421	metaphano
750	1.2	565	1	MAOX	SCHPO	O84290	chlamydia t	823	6	1.2	644	1	RNR	HELPU	Q9u1t5	schizosacch
751	1.2	565	1	MOBA	BIFLO	P40375	schizosacch	824	6	1.2	648	1	APPC	DICDI	Q9zjx9	helicobacte
752	1.2	565	1	MOBA	BIFLO	O8gn31	biflobacte	825	6	1.2	648	1	PRIA	RICPR	P34059	dictyosteli
753	1.2	566	1	DCPY	HAELN	P45286	haemophilus	826	6	1.2	648	1	RCOI	MOUSE	Q9z219	mus musculus
754	1.2	566	1	PAP	SCHPO	P87208	ameiobacilla	827	6	1.2	649	1	RCOI	MOUSE	P46063	homo sapien
755	1.2	567	1	GFV	RAT	Q10295	echizosacch	828	6	1.2	652	1	NAK1	SCHPO	O75011	schizosacch
756	1.2	568	1	OXO	OXAPO	O08770	rattus norv	829	6	1.2	652	1	NOSS2	PARDE	Q51705	paracoccus
757	1.2	569	1	UREI	SYNY3	P40149	oxalobacter	830	6	1.2	653	1	API	ACHLY	P15636	achromobact
758	1.2	570	1	SMA4	CAREL	P45897	caenorhabdi	831	6	1.2	653	1	PABP	SCHPO	P31209	schizosacch
759	1.2	571	1	DISJ	BORTA	P30431	bothrops ja	832	6	1.2	655	1	MLAI	MOUSE	P45700	mus musculus
760	1.2	571	1	PAL4	POPKI	Q40910	populus kit	833	6	1.2	660	1	DNL2	BORBU	O51502	borrelia bu
761	1.2	572	1	MAOX	HUMAN	P48163	homo sapien	834	6	1.2	660	1	SYR	HUMAN	P54136	homo sapien
762	1.2	572	1	MOES	LYTVA	P53962	lytechinus	835	6	1.2	660	1	SYR	MOUSE	Q9d019	mus musculus
763	1.2	572	1	UREI	STRSL	P50047	strepitococc	836	6	1.2	660	1	YHL1	EBV	P03181	epstein-bar



837	6	1.2	661	1	SYR_CRILLO	P37880	cricetus	910	6	1.2	766	1	BCSB_SALTY	Q931n1	salmonella
838	6	1.2	662	1	ITR_HUMAN	P42701	homo sapien	911	6	1.2	766	1	GAPI_SCHPO	P33377	schizosach
839	6	1.2	662	1	LOX2_RAT	Q02759	rattus norv	912	6	1.2	767	1	PSAA_SYNPW	Q92090	synchococc
840	6	1.2	662	1	LOXL_MOUSE	P39654	mus musculus	913	6	1.2	773	1	ITB1_BOVIN	P53712	bos taurus
841	6	1.2	662	1	NX2B_RAT	Q63376	rattus norv	914	6	1.2	773	1	PSAA_PROMA	Q91494	prochlocco
842	6	1.2	663	1	DN1J_CHLTR	Q84148	chlamydia t	915	6	1.2	774	1	RRP3_INCB	P21770	influenza c
843	6	1.2	666	1	NX28_HUMAN	P58401	homo sapien	916	6	1.2	775	1	TREY_ARTSQ	Q44315	artrobacte
844	6	1.2	667	1	M121_DROME	P53624	drosophila	917	6	1.2	776	1	TFR1_CHICK	Q09097	gallus gall
845	6	1.2	667	1	PALY_HILAN	Q04058	heliandus	918	6	1.2	780	1	VPA_ROTET	Q02945	equine rota
846	6	1.2	668	1	FE08_METJA	Q57868	methanococc	919	6	1.2	781	1	SAV_SULAC	Q00290	sulfolobus
847	6	1.2	668	1	PAB5_ARATH	Q05196	arabidopsis	920	6	1.2	781	1	MERV_HUMAN	Q15553	homo sapien
848	6	1.2	668	1	PBPC_BACSU	P42971	baclillus su	921	6	1.2	791	1	PSA_ECOLI	P23338	escherichia
849	6	1.2	669	1	AMY_ALTRA	P29957	altermonas	922	6	1.2	793	1	FUR1_MOUSE	P23188	mus musculus
850	6	1.2	669	1	NUSM_ARATH	P23888	arabidopsis	923	6	1.2	795	1	SYFB_VIBVU	Q8d439	vibrio vuln
851	6	1.2	670	1	NUSM_ONBNE	P10330	oenochera b	924	6	1.2	797	1	PBPA_NEIGO	Q05131	neisseria g
852	6	1.2	670	1	NUSM_WHEAT	Q37680	triticum ae	925	6	1.2	797	1	SHK3_HUMAN	Q9b3b0	homo sapien
853	6	1.2	671	1	2282_HUMAN	Q9u4v7	homo sapien	926	6	1.2	798	1	ITB1_PELCA	P53713	felis silve
854	6	1.2	675	1	PPK_HELPY	Q92m10	helicobacte	927	6	1.2	798	1	ITB1_HUMAN	P05556	homo sapien
855	6	1.2	675	1	PPK_HELPY	Q25654	helicobacte	928	6	1.2	798	1	ITB1_MOUSE	P09055	mus musculus
856	6	1.2	675	1	UVRB_NEIGO	Q50939	neisseria g	929	6	1.2	798	1	PBPA_NEICI	Q87626	neisseria f
857	6	1.2	675	1	UVRB_NEIMA	P56996	neisseria m	930	6	1.2	798	1	PBPA_NEIFL	Q87529	neisseria l
858	6	1.2	675	1	UVRB_NEIMA	Q33395	neisseria m	931	6	1.2	798	1	PBPA_NEIMA	Q05194	neisseria m
859	6	1.2	679	1	HS82_AJECA	P33125	ajellomyces	932	6	1.2	799	1	ITB1_RAT	P49134	rattus norv
860	6	1.2	679	1	TKT1_YEAST	P23254	saccharomyc	933	6	1.2	803	1	GYRB_ECOLI	Q60008	salmonella
861	6	1.2	679	1	TKT1_YEAST	Q42675	craterostig	934	6	1.2	803	1	GYRB_SALTY	Q15320	homo sapien
862	6	1.2	680	1	KALM_HUMAN	P23352	homo sapien	935	6	1.2	804	1	MEAG_HUMAN	P24932	human adeno
863	6	1.2	681	1	NUSC_SYNY3	Q55429	synchocyst	936	6	1.2	805	1	L100_ADE02	P24932	human adeno
864	6	1.2	681	1	YADD_SCHPO	Q09838	schizosach	937	6	1.2	806	1	NIUG_MYCTU	P24933	mycobacteri
865	6	1.2	692	1	BOMD_XENLA	P79944	xenopus lae	938	6	1.2	807	1	L100_ADE05	P24933	mycobacteri
866	6	1.2	702	1	AMYG_CLOSP	P29761	clostridium	939	6	1.2	807	1	L100_ADE05	P24933	mycobacteri
867	6	1.2	703	1	CDGT_BACS2	P31746	baclillus sp	940	6	1.2	808	1	PLSB_VIBPA	Q874n0	vibrio para
868	6	1.2	703	1	SYM_CHUTE	P59077	chlorobium	941	6	1.2	810	1	K105_YEAST	Q40594	saccharomyc
869	6	1.2	705	1	MMUC_STRCO	O88022	streptomyces	942	6	1.2	814	1	METE_CHLRE	Q39586	chlamydomon
870	6	1.2	705	1	MX_CHICK	Q90597	gallus gall	943	6	1.2	824	1	J1P2_HUMAN	Q13387	homo sapien
871	6	1.2	705	1	YKX2_YEAST	P35950	saccharomyc	944	6	1.2	829	1	J1P2_HABIN	P44323	haemophilus
872	6	1.2	706	1	ULI7_HSVB	P28950	equine hefp	945	6	1.2	829	1	VIRA_AGR19	P10799	agrobacteri
873	6	1.2	707	1	SFPQ_HUMAN	P23246	homo sapien	946	6	1.2	829	1	VIRA_AGR19	P10799	agrobacteri
874	6	1.2	708	1	PAL2_IPOBA	Q42858	ipomoea bat	947	6	1.2	832	1	YFC4_YEAST	P43752	saccharomyc
875	6	1.2	712	1	PAL2_TOBAC	P35513	nicotiana t	948	6	1.2	833	1	J1P2_PASMU	P57573	saccharomyc
876	6	1.2	712	1	PAL3_TOBAC	P45733	nicotiana t	949	6	1.2	835	1	NIUG2_HUMAN	Q8n742	homo sapien
877	6	1.2	713	1	PAL1_SOYBN	P27991	glycine max	950	6	1.2	835	1	TR1B_HUMAN	Q13663	homo sapien
878	6	1.2	713	1	SIR_CAEEL	Q19625	caenorhabd	951	6	1.2	836	1	NIUG2_RAT	Q62888	rattus norv
879	6	1.2	715	1	PAL1_TOBAC	P28672	nicotiana t	952	6	1.2	843	1	CO7_HUMAN	P10643	homo sapien
880	6	1.2	718	1	PAL2_CICAR	Q9smk9	cicer ariet	953	6	1.2	847	1	SYA_HELPJ	Q921y5	helicobacte
881	6	1.2	720	1	NUP2_YEAST	P32499	saccharomyc	954	6	1.2	847	1	YFC4_YEAST	P43752	saccharomyc
882	6	1.2	720	1	PAL1_SOLUTU	P31425	solanum tub	955	6	1.2	851	1	MCEL_ROMPY	Q93584	foxiopox vir
883	6	1.2	721	1	PAL5_LYCCE	P26600	lycopersico	956	6	1.2	853	1	PBPA_HABIN	P31776	h penicilli
884	6	1.2	722	1	PAL1_CITLI	Q42667	citrus limo	957	6	1.2	861	1	PQ58_CAEEL	P34552	caenorhabd
885	6	1.2	723	1	PAL1_PEA	Q01861	pisum sativ	958	6	1.2	862	1	SLA2_BACAN	P94517	baclillus an
886	6	1.2	724	1	PAL2_PEA	Q04593	pisum sativ	959	6	1.2	867	1	POL_MPMV	P04625	simian mabo
887	6	1.2	725	1	PALY_MEDSA	P27990	medicago sa	960	6	1.2	867	1	POL_SRV1	P04625	simian retr
888	6	1.2	725	1	PALY_TRISU	P45734	trifolium s	961	6	1.2	867	1	POL_SRV2	P51517	simian retr
889	6	1.2	727	1	IF2M_BOVIN	P46198	bos taurus	962	6	1.2	868	1	SRYC_DROME	P16519	drosophila
890	6	1.2	728	1	CATB_ASFPU	Q92405	aspergillus	963	6	1.2	870	1	BCA1_HUMAN	P56945	homo sapien
891	6	1.2	730	1	YP65_CAEEL	Q09214	caenorhabd	964	6	1.2	870	1	COPG_CAEEL	Q22498	caenorhabd
892	6	1.2	730	1	CATR_ASFPG	P55303	aspergillus	965	6	1.2	870	1	DYN2_HUMAN	P507570	homo sapien
893	6	1.2	735	1	DHB4_MOUSE	P51660	mus musculus	966	6	1.2	870	1	DYN2_MOUSE	P39954	mus musculus
894	6	1.2	738	1	YO13_BRL2	P42548	bacterioph	967	6	1.2	870	1	DYN2_RAT	P39954	rattus norv
895	6	1.2	742	1	DHET_ACEAC	P18278	acetobacter	968	6	1.2	870	1	POL_JSRV	P31623	sheep pulmo
896	6	1.2	743	1	KPC2_APLCA	Q16975	aplysia cal	969	6	1.2	874	1	BCA1_MOUSE	Q61140	mus musculus
897	6	1.2	743	1	YN02_CAEEL	Q03606	caenorhabd	970	6	1.2	878	1	SECA_ANTSP	Q61140	mus musculus
898	6	1.2	745	1	VAT_THEAC	Q05209	thermoplas	971	6	1.2	878	1	SYL_TREPA	Q83395	treponema p
899	6	1.2	752	1	GC20_YEAST	P53535	saccharomyc	972	6	1.2	887	1	MCW2_DROME	P49735	drosophila
900	6	1.2	752	1	XDHA_ECOLI	O8x6c7	escherichia	973	6	1.2	888	1	POL_SRYVH	P03364	squitrrel mo
901	6	1.2	752	1	XDHA_ECOLI	Q46799	escherichia	974	6	1.2	892	1	YLB3_CAEEL	P46578	caenorhabd
902	6	1.2	758	1	GCR_ONOMY	P49843	oncorhynch	975	6	1.2	895	1	SECA_CYACA	Q19911	cyatidium c
903	6	1.2	763	1	TSHR_BOVIN	Q27987	bos taurus	976	6	1.2	897	1	CHIA_ECOLI	P13365	escherichia
904	6	1.2	764	1	TSHR_CANFA	P14763	canis famli	977	6	1.2	899	1	POL_MMTVB	P03365	mouse mamma
905	6	1.2	764	1	TSHR_MOUSE	P47750	mus musculus	978	6	1.2	902	1	NRG4_TREPA	O83128	treponema p
906	6	1.2	764	1	TSHR_RAT	P21463	rattus norv	979	6	1.2	903	1	NRG4_HUMAN	Q14934	homo sapien
907	6	1.2	764	1	TSHR_SHEEP	P56495	ovis aries	980	6	1.2	905	1	YB56_METJA	Q58556	methanococc
908	6	1.2	765	1	TREY_MCTU	Q10768	mycobacteri	981	6	1.2	905	1	HXA1_HAEIN	P44602	haemophilus
909	6	1.2	766	1	BCSB_SALTY	Q82290	salmonella	982	6	1.2	909	1	RPS2_ARATH	Q42484	arabidopsis

983	6	1.2	909	1	SBCC DEIRA	Q9t44 deiococcus
984	6	1.2	915	1	YLS4 CASEL	P4389 caenorhabdi
985	6	1.2	917	1	IL6B MOUSE	O00560 mus musculu
986	6	1.2	918	1	KRCM MOUSE	O62101 mus musculu
987	6	1.2	919	1	ATC1_HUMAN	P88199 homo sapien
988	6	1.2	919	1	ATC1_RAT	Q64566 ratu
989	6	1.2	919	1	SYNP_HUMAN	Q9y6h homo sapien
990	6	1.2	921	1	CA19_HUMAN	P0849 homo sapien
991	6	1.2	923	1	SECA_CAUCR	P38380 caulobacter
992	6	1.2	929	1	BLJ3_MOUSE	O9wv92 mus musculu
993	6	1.2	933	1	PLD2_MOUSE	P97813 mus musculu
994	6	1.2	933	1	PLD2_RAT	P70496 ratu
995	6	1.2	934	1	CITC_RAT	P27653 r c-1-tera
996	6	1.2	940	1	SEAI_MYCLE	P57956 mycobacteri
997	6	1.2	943	1	LBPA_NEIMA	O06379 neisseria m
998	6	1.2	944	1	LBPA_NEIMA	Q9tk4 neisseria m
999	6	1.2	948	1	SECA_MYCBO	P71499 mycobacteri
1000	6	1.2	949	1	SEAI_MYCTU	O05885 mycobacteri

## ALIGNMENTS

RESULT 1

ANIA_NEIGO	STANDARD;	PRT;	392 AA.
ID ANIA_NEIGO			
AC 002219;			
DT 01-JUN-1994 (Rel. 29, Created)			
DT 01-JUN-1994 (Rel. 29, Last sequence update)			
DT 15-SEP-2003 (Rel. 42, Last annotation update)			
DE Major outer membrane protein Pan 1 precursor.			
GN ANIA.			
OS Neisseria gonorrhoeae.			
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;			
OC Neisseriaceae; Neisseria.			
OX NCBI_TaxID=485;			
ON [1]			
RP SEQUENCE FROM N.A.			
RC STRAIN=RI0;			
RX MEDLINE=93014187; PubMed=1383156;			
RA Hoehn G.T., Clark V.L.			
RT "Isolation and nucleotide sequence of the gene (ania) encoding the			
RT gonorrhoeae."			
RL Infect. Immun. 60:4695-4703 (1992).			
RN [2]			
RP SEQUENCE FROM N.A.			
RC STRAIN=ATCC 33084 / F62;			
RX MEDLINE=93014188; PubMed=1398981;			
RA Hoehn G.T., Clark V.L.			
RT "The major anaerobically induced outer membrane protein of Neisseria			
RT gonorrhoeae, Pan 1, is a lipoprotein."			
RL Infect. Immun. 60:4704-4708 (1992).			
RN [2]			
RP SEQUENCE FROM N.A.			
RC STRAIN=ATCC 33084 / F62;			
RX MEDLINE=93014188; PubMed=1398981;			
RA Hoehn G.T., Clark V.L.			
RT "The major anaerobically induced outer membrane protein of Neisseria			
RT gonorrhoeae, Pan 1, is a lipoprotein."			
RL Infect. Immun. 60:4704-4708 (1992).			
RN [2]			
RP SEQUENCE FROM N.A.			
RC STRAIN=ATCC 33084 / F62;			
RX MEDLINE=93014188; PubMed=1398981;			
RA Hoehn G.T., Clark V.L.			
RT "The major anaerobically induced outer membrane protein of Neisseria			
RT gonorrhoeae, Pan 1, is a lipoprotein."			
RL Infect. Immun. 60:4704-4708 (1992).			
RN [2]			
RP SEQUENCE FROM N.A.			
RC STRAIN=ATCC 33084 / F62;			
RX MEDLINE=93014188; PubMed=1398981;			
RA Hoehn G.T., Clark V.L.			
RT "The major anaerobically induced outer membrane protein of Neisseria			
RT gonorrhoeae, Pan 1, is a lipoprotein."			
RL Infect. Immun. 60:4704-4708 (1992).			
RN [2]			
RP SEQUENCE FROM N.A.			
RC STRAIN=ATCC 33084 / F62;			
RX MEDLINE=93014188; PubMed=1398981;			
RA Hoehn G.T., Clark V.L.			
RT "The major anaerobically induced outer membrane protein of Neisseria			
RT gonorrhoeae, Pan 1, is a lipoprotein."			
RL Infect. Immun. 60:4704-4708 (1992).			
RN [2]			
RP SEQUENCE FROM N.A.			
RC STRAIN=ATCC 33084 / F62;			
RX MEDLINE=93014188; PubMed=1398981;			
RA Hoehn G.T., Clark V.L.			
RT "The major anaerobically induced outer membrane protein of Neisseria			
RT gonorrhoeae, Pan 1, is a lipoprotein."			
RL Infect. Immun. 60:4704-4708 (1992).			
RN [2]			
RP SEQUENCE FROM N.A.			
RC STRAIN=ATCC 33084 / F62;			
RX MEDLINE=93014188; PubMed=1398981;			
RA Hoehn G.T., Clark V.L.			
RT "The major anaerobically induced outer membrane protein of Neisseria			
RT gonorrhoeae, Pan 1, is a lipoprotein."			
RL Infect. Immun. 60:4704-4708 (1992).			
RN [2]			
RP SEQUENCE FROM N.A.			
RC STRAIN=ATCC 33084 / F62;			
RX MEDLINE=93014188; PubMed=1398981;			
RA Hoehn G.T., Clark V.L.			
RT "The major anaerobically induced outer membrane protein of Neisseria			
RT gonorrhoeae, Pan 1, is a lipoprotein."			
RL Infect. Immun. 60:4704-4708 (1992).			
RN [2]			
RP SEQUENCE FROM N.A.			
RC STRAIN=ATCC 33084 / F62;			
RX MEDLINE=93014188; PubMed=1398981;			
RA Hoehn G.T., Clark V.L.			
RT "The major anaerobically induced outer membrane protein of Neisseria			
RT gonorrhoeae, Pan 1, is a lipoprotein."			
RL Infect. Immun. 60:4704-4708 (1992).			
RN [2]			
RP SEQUENCE FROM N.A.			
RC STRAIN=ATCC 33084 / F62;			
RX MEDLINE=93014188; PubMed=1398981;			
RA Hoehn G.T., Clark V.L.			
RT "The major anaerobically induced outer membrane protein of Neisseria			
RT gonorrhoeae, Pan 1, is a lipoprotein."			
RL Infect. Immun. 60:4704-4708 (1992).			
RN [2]			
RP SEQUENCE FROM N.A.			
RC STRAIN=ATCC 33084 / F62;			
RX MEDLINE=93014188; PubMed=1398981;			
RA Hoehn G.T., Clark V.L.			
RT "The major anaerobically induced outer membrane protein of Neisseria			
RT gonorrhoeae, Pan 1, is a lipoprotein."			
RL Infect. Immun. 60:4704-4708 (1992).			
RN [2]			
RP SEQUENCE FROM N.A.			
RC STRAIN=ATCC 33084 / F62;			
RX MEDLINE=93014188; PubMed=1398981;			
RA Hoehn G.T., Clark V.L.			
RT "The major anaerobically induced outer membrane protein of Neisseria			
RT gonorrhoeae, Pan 1, is a lipoprotein."			
RL Infect. Immun. 60:4704-4708 (1992).			
RN [2]			
RP SEQUENCE FROM N.A.			
RC STRAIN=ATCC 33084 / F62;			
RX MEDLINE=93014188; PubMed=1398981;			
RA Hoehn G.T., Clark V.L.			
RT "The major anaerobically induced outer membrane protein of Neisseria			
RT gonorrhoeae, Pan 1, is a lipoprotein."			
RL Infect. Immun. 60:4704-4708 (1992).			
RN [2]			
RP SEQUENCE FROM N.A.			
RC STRAIN=ATCC 33084 / F62;			
RX MEDLINE=93014188; PubMed=1398981;			
RA Hoehn G.T., Clark V.L.			
RT "The major anaerobically induced outer membrane protein of Neisseria			
RT gonorrhoeae, Pan 1, is a lipoprotein."			
RL Infect. Immun. 60:4704-4708 (1992).			
RN [2]			
RP SEQUENCE FROM N.A.			
RC STRAIN=ATCC 33084 / F62;			
RX MEDLINE=93014188; PubMed=1398981;			
RA Hoehn G.T., Clark V.L.			
RT "The major anaerobically induced outer membrane protein of Neisseria			
RT gonorrhoeae, Pan 1, is a lipoprotein."			
RL Infect. Immun. 60:4704-4708 (1992).			
RN [2]			
RP SEQUENCE FROM N.A.			
RC STRAIN=ATCC 33084 / F62;			
RX MEDLINE=93014188; PubMed=1398981;			
RA Hoehn G.T., Clark V.L.			
RT "The major anaerobically induced outer membrane protein of Neisseria			
RT gonorrhoeae, Pan 1, is a lipoprotein."			
RL Infect. Immun. 60:4704-4708 (1992).			
RN [2]			
RP SEQUENCE FROM N.A.			
RC STRAIN=ATCC 33084 / F62;			
RX MEDLINE=93014188; PubMed=1398981;			
RA Hoehn G.T., Clark V.L.			
RT "The major anaerobically induced outer membrane protein of Neisseria			
RT gonorrhoeae, Pan 1, is a lipoprotein."			
RL Infect. Immun. 60:4704-4708 (1992).			
RN [2]			
RP SEQUENCE FROM N.A.			
RC STRAIN=ATCC 33084 / F62;			
RX MEDLINE=93014188; PubMed=1398981;			
RA Hoehn G.T., Clark V.L.			
RT "The major anaerobically induced outer membrane protein of Neisseria			
RT gonorrhoeae, Pan 1, is a lipoprotein."			
RL Infect. Immun. 60:4704-4708 (1992).			
RN [2]			
RP SEQUENCE FROM N.A.			
RC STRAIN=ATCC 33084 / F62;			
RX MEDLINE=93014188; PubMed=1398981;			
RA Hoehn G.T., Clark V.L.			
RT "The major anaerobically induced outer membrane protein of Neisseria			
RT gonorrhoeae, Pan 1, is a lipoprotein."			
RL Infect. Immun. 60:4704-4708 (1992).			
RN [2]			
RP SEQUENCE FROM N.A.			
RC STRAIN=ATCC 33084 / F62;			
RX MEDLINE=93014188; PubMed=1398981;			
RA Hoehn G.T., Clark V.L.			
RT "The major anaerobically induced outer membrane protein of Neisseria			
RT gonorrhoeae, Pan 1, is a lipoprotein."			
RL Infect. Immun. 60:4704-4708 (1992).			
RN [2]			
RP SEQUENCE FROM N.A.			
RC STRAIN=ATCC 33084 / F62;			
RX MEDLINE=93014188; PubMed=1398981;			
RA Hoehn G.T., Clark V.L.			
RT "The major anaerobically induced outer membrane protein of Neisseria			
RT gonorrhoeae, Pan 1, is a lipoprotein."			
RL Infect. Immun. 60:4704-4708 (1992).			
RN [2]			
RP SEQUENCE FROM N.A.			
RC STRAIN=ATCC 33084 / F62;			
RX MEDLINE=93014188; PubMed=1398981;			
RA Hoehn G.T., Clark V.L.			
RT "The major anaerobically induced outer membrane protein of Neisseria			
RT gonorrhoeae, Pan 1, is a lipoprotein."			
RL Infect. Immun. 60:4704-4708 (1992).			
RN [2]			
RP SEQUENCE FROM N.A.			
RC STRAIN=ATCC 33084 / F62;			
RX MEDLINE=93014188; PubMed=1398981;			
RA Hoehn G.T., Clark V.L.			
RT "The major anaerobically induced outer membrane protein of Neisseria			
RT gonorrhoeae, Pan 1, is a lipoprotein."			
RL Infect. Immun. 60:4704-4708 (1992).			
RN [2]			
RP SEQUENCE FROM N.A.			
RC STRAIN=ATCC 33084 / F62;			
RX MEDLINE=93014188; PubMed=1398981;			
RA Hoehn G.T., Clark V.L.			
RT "The major anaerobically induced outer membrane protein of Neisseria			
RT gonorrhoeae, Pan 1, is a lipoprotein."			
RL Infect. Immun. 60:4704-4708 (1992).			
RN [2]			
RP SEQUENCE FROM N.A.			
RC STRAIN=ATCC 33084 / F62;			
RX MEDLINE=93014188; PubMed=1398981;			
RA Hoehn G.T., Clark V.L.			
RT "The major anaerobically induced outer membrane protein of Neisseria			
RT gonorrhoeae, Pan 1, is a lipoprotein."			
RL Infect. Immun. 60:4704-4708 (1992).			
RN [2]			
RP SEQUENCE FROM N.A.			
RC STRAIN=ATCC 33084 / F62;			
RX MEDLINE=93014188; PubMed=1398981;			
RA Hoehn G.T., Clark V.L.			
RT "The major anaerobically induced outer membrane protein of Neisseria			
RT gonorrhoeae, Pan 1, is a lipoprotein."			
RL Infect. Immun. 60:4704-4708 (1992).			
RN [2]			
RP SEQUENCE FROM N.A.			
RC STRAIN=ATCC 33084 / F62;			
RX MEDLINE=93014188; PubMed=1398981;			
RA Hoehn G.T., Clark V.L.			
RT "The major anaerobically induced outer membrane protein of Neisseria			
RT gonorrhoeae, Pan 1, is a lipoprotein."			
RL Infect. Immun. 60:4704-4708 (1992).			
RN [2]			
RP SEQUENCE FROM N.A.			
RC STRAIN=ATCC 33084 / F62;			
RX MEDLINE=93014188; PubMed=1398981;			
RA Hoehn G.T., Clark V.L.			
RT "The major anaerobically induced outer membrane protein of Neisseria			
RT gonorrhoeae, Pan 1, is a lipoprotein."			
RL Infect. Immun. 60:4704-4708 (1992).			
RN [2]			
RP SEQUENCE FROM N.A.			
RC STRAIN=ATCC 33084 / F62;			
RX MEDLINE=93014188; PubMed=1398981;			
RA Hoehn G.T., Clark V.L.			
RT "The major anaerobically induced outer membrane protein of Neisseria			
RT gonorrhoeae, Pan 1, is a lipoprotein."			
RL Infect. Immun. 60:4704-4708 (1992).			
RN [2]			
RP SEQUENCE FROM N.A.			
RC STRAIN=ATCC 33084 / F62;			
RX MEDLINE=93014188; PubMed=1398981;			
RA Hoehn G.T., Clark V.L.			
RT "The major anaerobically induced outer membrane protein of Neisseria			
RT gonorrhoeae, Pan 1, is a lipoprotein."			
RL Infect. Immun. 60:4704-4708 (1992).			

```
DR InterPro: IPR001287; CuNO2_reductase.
DR Pfam: PF00394; Cu-oxidase; 2.
DR PRINTS: PR00695; CONO2RDTASE.
KW Oxidoreductase; Copper; Metal-binding; Flavoprotein; FAD;
KW Nitrate assimilation; Repeat; Periplasmic; 3d-structure.
FT DOMAIN 1 165 PLASTOCYANIN-LIKE 1.
FT METAL 166 330 PLASTOCYANIN-LIKE 2.
FT METAL 85 85 COPPER (TYPE 1).
FT METAL 90 90 COPPER (TYPE 2).
FT METAL 125 125 COPPER (TYPE 2).
FT METAL 126 126 COPPER (TYPE 1).
FT METAL 135 135 COPPER (TYPE 1).
FT METAL 140 140 COPPER (TYPE 1).
FT METAL 296 296 COPPER (TYPE 2).
SQ SEQUENCE 330 AA; 34413 MW; E2C38C3A2CEBFCE8 CRC64;

Query Match 2.4%; Score 12; DB 1; Length 330;
Best Local Similarity 100.0%; Pred. No. 0.0003;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 137 MPNVDPHATG 148
DB 83 MPNVDPHATG 94

RESULT 3
NIR_PSECL STANDARD; PRT; 363 AA.
AC 006006;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Copper-containing nitrite reductase precursor (EC 1.7.2.1) (Cu-NIR).
GN NIKR.
OS Pseudomonas chlororaphis (Pseudomonas aureofaciens).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCB1_TaxID=333;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 25-41.
RC STRAIN=ATCC 13985;
RX MEDLINE=93356602; PubMed=8352648;
RA Glockner A.B., Juenget A., Zumft W.G.;
RT "Copper-containing nitrite reductase from Pseudomonas aureofaciens is functional in a mutationally cytochrome cdl-free background (Nirs-)"
RT of Pseudomonas stutzeri.";
RL Arch. Microbiol. 160:18-26(1993).
CC -1- CATALYTIC ACTIVITY: Nitric oxide + H(2)O + ferricytochrome c =
CC nitrite + ferrocyclochrome c.
CC -1- COPFACTOR: CONTAINS ONE ATOM OF TYPE I COPPER AND ONE ATOM OF TYPE
CC II COPPER; FAD. TYPE I COPPER IS BOUND WITHIN A SINGLE MONOMER,
CC WHILE TYPE 2 COPPER IS HELD BY RESIDUES FROM EACH OF TWO MONOMERS
CC OF THE TRIMER. NITRITE IS BOUND TO THE TYPE 2 COPPER SITE.
CC PSEUDOMAZURIN IS THE PHYSIOLOGICAL ELECTRON DONOR FOR THE CU-NIR IN
CC VITRO.
CC -1- PATHWAY: Nitrate assimilation (denitrification).
CC -1- SUBUNIT: Homotrimer (By similarity).
CC -1- SUBCELLULAR LOCATION: Periplasmic.
CC -1- DOMAIN: THE TYPE I COPPER SITE IN NIR PLAYS A CRUCIAL ROLE FOR
CC ELECTRON TRANSFER FROM PSEUDOMAZURIN TO THE TYPE II COPPER SITE
CC OF NIR, WHICH COMPRISES THE CATALYTIC CENTER OF NIR FOR THE
CC REDUCTION OF NITRITE.
CC -1- SIMILARITY: Contains 2 plastocyanin-like domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z21945; CAA79939.1; .
```

```
DR PIR; S32112; S32112.
DR HSSP; P38501; 1A58.
DR InterPro: IPR001117; Cu-oxidase.
DR InterPro: IPR001287; CuNO2_reductase.
DR Pfam: PF00394; Cu-oxidase; 2.
DR PRINTS: PR00695; CONO2RDTASE.
KW Oxidoreductase; Copper; Metal-binding; Flavoprotein; FAD;
KW Nitrate assimilation; Repeat; Periplasmic; Signal.
FT SIGNAL 1 24
FT CHAIN 25 363 COPPER-CONTAINING NITRITE REDUCTASE.
FT DOMAIN 25 193 PLASTOCYANIN-LIKE 1.
FT METAL 194 363 PLASTOCYANIN-LIKE 2.
FT METAL 113 113 COPPER (TYPE 1) (BY SIMILARITY).
FT METAL 118 118 COPPER (TYPE 2) (BY SIMILARITY).
FT METAL 153 153 COPPER (TYPE 2) (BY SIMILARITY).
FT METAL 154 154 COPPER (TYPE 1) (BY SIMILARITY).
FT METAL 163 163 COPPER (TYPE 1) (BY SIMILARITY).
FT METAL 168 168 COPPER (TYPE 1) (BY SIMILARITY).
FT METAL 324 324 COPPER (TYPE 2) (BY SIMILARITY).
SQ SEQUENCE 363 AA; 39248 MW; 50DB60CC4DC3E00 CRC64;

Query Match 2.4%; Score 12; DB 1; Length 363;
Best Local Similarity 100.0%; Pred. No. 0.00032;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 137 MPNVDPHATG 148
DB 111 MPNVDPHATG 122

RESULT 4
NIR_RHIHE STANDARD; PRT; 377 AA.
AC 060214;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Copper-containing nitrite reductase precursor (EC 1.7.2.1) (Cu-NIR).
GN NIKR.
OS Rhizobium hedyarai.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium/Agrobacterium group; Rhizobium.
OX NCB1_TaxID=50338;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HCNT1;
RX MEDLINE=97055705; PubMed=8699992;
RA Toffanin A., Wu Q., Maskus M., Casella S., Abruna H.D.,
RA Shapleigh J.P.;
RT "Characterization of the gene encoding nitrite reductase and the
RT physiological consequences of its expression in the nondenitrifying
RT Rhizobium hedyarai strain HCNT1.";
RL Appl. Environ. Microbiol. 62:4019-4025(1996).
CC -1- CATALYTIC ACTIVITY: Nitric oxide + H(2)O + ferricytochrome c =
CC nitrite + ferrocyclochrome c.
CC -1- COPFACTOR: CONTAINS ONE ATOM OF TYPE I COPPER AND ONE ATOM OF TYPE
CC II COPPER; FAD. TYPE I COPPER IS BOUND WITHIN A SINGLE MONOMER,
CC WHILE TYPE 2 COPPER IS HELD BY RESIDUES FROM EACH OF TWO MONOMERS
CC OF THE TRIMER. NITRITE IS BOUND TO THE TYPE 2 COPPER SITE.
CC PSEUDOMAZURIN IS THE PHYSIOLOGICAL ELECTRON DONOR FOR THE CU-NIR IN
CC VITRO (BY SIMILARITY).
CC -1- PATHWAY: Nitrate assimilation (denitrification).
CC -1- SUBUNIT: Homotrimer (By similarity).
CC -1- SUBCELLULAR LOCATION: Periplasmic (By similarity).
CC -1- DOMAIN: THE TYPE I COPPER SITE IN NIR PLAYS A CRUCIAL ROLE FOR
CC ELECTRON TRANSFER FROM PSEUDOMAZURIN TO THE TYPE II COPPER SITE
CC OF NIR, WHICH COMPRISES THE CATALYTIC CENTER OF NIR FOR THE
CC REDUCTION OF NITRITE.
CC -1- SIMILARITY: Contains 2 plastocyanin-like domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
```

CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; U65658; AAB05880.1; -.  
 DR HSP; P38501; 1AS7.  
 DR InterPro: IPR001117; Cu-oxidase.  
 DR InterPro: IPR001287; CuNO2\_reductase.  
 DR InterPro: IPR006311; Tat.  
 DR Pfam; PF00394; Cu-oxidase; 2.  
 DR PRINTS; PR00695; CUNO2RDTASE.  
 DR TIGRFAMs; TIGR01409; Tat\_signal\_seq; 1.  
 KM Oxidoreductase; Copper; Metal-binding; Flavoprotein; FAD;  
 KM Nitrate assimilation; Repeat; Periplasmic; Signal.  
 FT SIGNAL 1 46  
 FT CHAIN 1 46  
 FT DOMAIN 47 377  
 FT DOMAIN 99 194  
 FT DOMAIN 259 360  
 FT METAL 132 132  
 FT METAL 132 132  
 FT METAL 137 137  
 FT METAL 172 172  
 FT METAL 173 173  
 FT METAL 182 182  
 FT METAL 187 187  
 FT METAL 343 343  
 SQ SEQUENCE 377 AA; 40719 MW; C7AFBE9AE639E288 CRC64;

Query Match 2.0%; Score 10; DB 1; Length 377;  
 Best Local Similarity 100.0%; Pred. No. 0.039;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 252 VGALTGENAL 261  
 DB 261 VGALTGENAL 270

RESULT 5  
 RCSF\_ECOLI  
 ID RCSF\_ECOLI STANDARD; PRT; 134 AA.  
 AC P28633;  
 DT 01-DEC-1992 (Rel. 24, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Protein rcsf.  
 GN RCSF OR B0196 OR C0237.  
 OS Escherichia coli, and  
 OS Escherichia coli O6.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.  
 OX NCBI\_TaxID=562, 217992;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=93094132; PubMed=1459951;  
 RA Gervais F.G., Drapeau G.R.;  
 RT "Identification, cloning, and characterization of rcsf, a new  
 RT regulator gene for exopolysaccharide synthesis that suppresses the  
 RT division mutation ftsZ84 in Escherichia coli K-12.";  
 RL J. Bacteriol. 174:8016-8022(1992).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RA Miyamoto K.;  
 RL Submitted (APR-1993) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MG1655;  
 RX MEDLINE=97426617; PubMed=9278503;  
 RA Blatter F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Coliado-Vuades J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.;

RT "The complete genome sequence of Escherichia coli K-12.";  
 RL Science 277:1453-1474(1997).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / W3110;  
 RA Takemoto K., Mori H., Murayama N., Kataoka K., Yano M., Itoh T.,  
 RA Yamamoto Y., Inokuchi H., Miki T., Hatada E., Fukuda R., Ichihara S.,  
 RA Mizuno T., Makino K., Nakata A., Yura T., Samped G., Mizobuchi K.;  
 RT "Systematic sequencing of the Escherichia coli genome: analysis of the  
 RT 4.0-6.0 min (189,987-281,416bp) region";  
 RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RA Schramm S., Duncan M., Allen E., Araujo R., Aparicio A., Chung E.,  
 RA Davis K., Federspiel N., Hyman R., Kalman S., Komp C., Kundi O.,  
 RA Laskari D., Lew H., Lin D., Namath A., Oefner P., Roberts D.,  
 RA Davis R.W.;  
 RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=06:H1 / CPT073 / ATCC 700928;  
 RX MEDLINE=22388234; PubMed=12471157;  
 RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,  
 RA Raske D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,  
 RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,  
 RA Mobley H.L.T., Donnenberg M.S., Blatter F.R.;  
 RT "Extensive mosaic structure revealed by the complete genome sequence  
 RT of uropathogenic Escherichia coli.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).  
 CC -I- FUNCTION: Stimulates colanic acid capsule synthesis. May  
 CC promote the phosphorylation of rcsB.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; L04474; AAA24508.1; -.  
 DR EMBL; D15061; BAA03656.1; -.  
 DR EMBL; AE000128; AAC73307.1; -.  
 DR EMBL; D83536; BAA7873.1; -.  
 DR EMBL; U70214; AAB08624.1; -.  
 DR EMBL; AE016755; AAN78729.1; -.  
 DR PIR; D64744; D64744.  
 DR EcoGene; Egl1502; rcsf.  
 KM Bacterial capsule; Complete proteome.  
 FT CONFLICT 37 38  
 FT CONFLICT 46 46 MISSING (IN REF. 1 AND 4).  
 SQ SEQUENCE 134 AA; 14163 MW; 824F8251C07BB41 CRC64;

Query Match 1.6%; Score 8; DB 1; Length 134;  
 Best Local Similarity 100.0%; Pred. No. 1.9;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 ALMLSGCS 24  
 DB 10 ALMLSGCS 17

RESULT 6  
 ID\_PORA\_METUA STANDARD; PRT; 389 AA.  
 AC O57715;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Pyruvate synthase subunit porA (EC 1.2.7.1) (Pyruvate oxidoreductase  
 DE alpha chain) (POR) (Pyruvic-ferredoxin oxidoreductase alpha subunit).  
 GN PORA OR MJ0267.  
 OS Methanococcus jannaschii.

OC Archaea; Euryarchaeota; Methanococci; Methanococcales;  
OC Methanocaldococcaceae; Methanocaldococcus.  
OX NCBI\_TaxID=2190;  
[1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=JUL-1 / DSM 2661 / ATCC 43067;  
RA MEDLINE=96337999; PubMed=868087;  
RX Bult C.U., White O., Olsen G.J., Zhou L., Fleischmann R.D.,  
RA Sultun G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,  
RA Kierulff A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,  
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,  
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrman J.L., Nguyen D.,  
RA Uterback L.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,  
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,  
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;  
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus  
jannaschii";  
RU Science 273:1058-1073(1996).  
CC -1- CATALYTIC ACTIVITY: Pyruvate + CoA + oxidized ferredoxin = acetyl-  
CC -1- SUBUNIT: HETEROTETRAMER OF ONE ALPHA, ONE BETA, ONE DELTA AND ONE  
CC GAMMA CHAIN (BY SIMILARITY).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; U67482; AAB98254.1; -  
CC DR PIR; D64333; D64333.  
DR TIGR; MJO267; -  
DR InterPro: IPR002880; POR\_N.  
DR Pfam; PF01855; POR\_N; 1.  
KW Oxidoreductase; Complete proteome.  
SQ SEQUENCE 389 AA; 42971 MW; 2CA3DDE3EB0F5A9 CRC64;  
  
Query Match 1.6%; Score 8; DB 1; Length 389;  
Best Local Similarity 100.0%; Pred. No. 4.8;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0.  
  
QY 340 FNKGALGI 347  
Db 332 FNKGALGI 339  
  
RESULT 7  
TIG-THETN STANDARD; PRT; 432 AA.  
AC O8RC26;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DE Trieger factor (TF).  
GN TIG OR TTE0624.  
OS Thermoaerobacter tengcongensis.  
OC Bacteria; Firmicutes; Clostridia; Thermoaerobacteriales;  
OC Thermoaerobacteriaceae; Thermoaerobacter.  
OX NCBI\_TaxID=119072;  
[1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MB4 / JCM 11007;  
RA MEDLINE=21992816; PubMed=11997336;  
RX Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,  
RA Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,  
RA Tan H., Chen R., Wang J., Yu J., Yang H.;  
RT "A complete sequence of T. tengcongensis genome.";  
RU Genome Res. 12:689-700(2002).  
CC -1- FUNCTION: Involved in protein export. Acts as a chaperone by  
CC maintaining the newly synthesized protein in an open conformation  
CC (By similarity).

CC -1- SIMILARITY: BELONGS TO THE FKBP-TYPE PIPIASE FAMILY. TIG SUBFAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; AB013031; AM23893.1; -  
CC DR HAMAP; MF\_00303; -1- FKBP-PIPIASE.  
DR InterPro: IPR001179; FKBP-PIPIASE.  
DR InterPro: IPR005215; Tifg\_fac.  
DR TIGRPFAMs; TIGR00115; tifg\_1.  
DR PROSITE; PS00453; FKBP-PIPIASE\_1; FALSE\_NEG.  
DR PROSITE; PS00454; FKBP-PIPIASE\_2; FALSE\_NEG.  
DR PROSITE; PS00559; FKBP-PIPIASE\_3; 1.  
KW Cell division; Chaperone; Isomerase; Rotamase; Complete proteome.  
FT DOMAIN 163 248 PIPIASE, FKBP-TYPE.  
SQ SEQUENCE 432 AA; 49762 MW; F7DC7DC821072E15 CRC64;  
  
Query Match 1.6%; Score 8; DB 1; Length 432;  
Best Local Similarity 100.0%; Pred. No. 5.2;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0.  
  
QY 296 FEGKGEN 303  
Db 179 FEGKGEN 186  
  
RESULT 8  
SAP-CHICK  
ID -SAP-CHICK STANDARD; PRT; 518 AA.  
AC O13035;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Proactivator polypeptide precursor [Contains: Saposin A; Saposin B;  
DE Saposin C; Saposin D].  
GN PSAP.  
OS Gallus gallus (Chicken).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
OC Gallus.  
OX NCBI\_TaxID=9031;  
[1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 194-203.  
RC TISSUE=Brain, and Liver;  
RX MEDLINE=98129745; PubMed=9461526;  
RA Azuma N., Seo H.-C., Lie O., Fu O., Gould R.M., Hiraiwa M., Burt D.W.,  
RA Paton I.R., Morrice D.R., O'Brien J.S., Kishimoto Y.;  
RT "Cloning, expression and map assignment of chicken prosaposin.";  
RU Biochem. J. 330:321-327(1998).  
[2]  
RP SEQUENCE FROM N.A.  
RA Altman N., Horowitz M.;  
RU Submitted (NOV-1998) to the EMBL/Genbank/DBJ databases.  
CC -1- FUNCTION: THE LYSOSOMAL DEGRADATION OF SPHINGOLIPIDS TAKES PLACE  
CC BY THE SEQUENTIAL ACTION OF SPECIFIC HYDROLASES. SOME OF THESE  
CC ENZYMES REQUIRE SPECIFIC LOW-MOLECULAR MASS, NON-ENZYMIC PROTEINS:  
CC THE SPHINGOLIPIDS ACTIVATOR PROTEINS (COPROTEINS) (BY SIMILARITY).  
CC -1- FUNCTION: SAPOIN A AND SAPOIN C STIMULATE THE HYDROLYSIS OF  
CC GLUCOSYLCERAMIDE BY BETA-GLUCOSYLCERAMIDASE (EC 3.2.1.45) AND  
CC GALACTOSYLCERAMIDE BY BETA-GALACTOSYLCERAMIDASE (EC 3.2.1.46).  
CC SAPOIN-C APPARENTLY ACTS BY COMBINING WITH THE ENZYME AND ACIDIC  
CC LIPID TO FORM AN ACTIVATED COMPLEX, RATHER THAN BY SOLUBILIZING  
CC THE SUBSTRATE (BY SIMILARITY).  
CC -1- FUNCTION: SAPOIN B STIMULATES THE HYDROLYSIS OF GALACTO-  
CC CEREBROSIDE SULFATE BY ARYL SULFATASE A (EC 3.1.6.8), GM1  
CC GANGLIOSIDES BY BETA-GALACTOSIDASE (EC 3.2.1.23) AND  
CC GLOBOTRIASOLYCERAMIDE BY ALPHA-GALACTOSIDASE A (EC 3.2.1.22).  
CC SAPOIN-B FORMS A SOLUBILIZING COMPLEX WITH THE SUBSTRATES OF THE

```

CC SPHINGOLIPID HYDROLASES (BY SIMILARITY).
CC -1- FUNCTION: SAPOSIN D IS A SPECIFIC SPHINGOMYELIN PHOSPHODIESTERASE
CC ACTIVATOR (SC 3.1.4.12) (BY SIMILARITY).
CC -1- SUBUNIT: SAPOSIN B IS A HOMODIMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Lysosomal (By similarity).
CC -1- PFM: THIS PRECURSOR IS PROTEOLYTICALLY PROCESSED TO A SMALL
CC PEPTIDES, WHICH ARE SIMILAR TO EACH OTHER AND ARE SPHINGOLIPID
CC HYDROLASE ACTIVATOR PROTEINS (BY SIMILARITY).
CC -1- SIMILARITY: Contains 2 saposin A-type domains.
CC -1- SIMILARITY: Contains 4 saposin B-type domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AB003471; BAA19914.1; -
CC EMBL; AF108656; AAF05899.1; -
CC InterPro; IPR003119; Sapa.
CC InterPro; IPR000004; SapaB.
CC DR Pfam; PF02199; Sapa. 2.
CC DR Pfam; PF05184; Sapa_1; 4.
CC DR Pfam; PF03489; Sapa_2; 4.
CC DR ProDom; PD001732; SapaB sub. 3.
CC KM Signal; Glycoprotein; Lysosome; Sphingolipid metabolism; Repeat;
CC Gm2-gangliosidosis.
CC FT SIGNAL 1 17 POTENTIAL.
CC FT PROPEP 18 60
CC FT CHAIN 61 143 SAPOSIN A.
CC FT PROPEP 145 193
CC FT CHAIN 194 276 SAPOSIN B.
CC FT PROPEP 278 305
CC FT CHAIN 307 387 SAPOSIN C.
CC FT PROPEP 389 398
CC FT CHAIN 399 480
CC FT PROPEP 482 518 SAPOSIN D.
CC FT DOMAIN 22 55
CC FT DOMAIN 60 143 SAPOSIN-LIKE TYPE A 1.
CC FT DOMAIN 193 276 SAPOSIN-LIKE TYPE B 1.
CC FT DOMAIN 307 388 SAPOSIN-LIKE TYPE B 2.
CC FT DOMAIN 399 480 SAPOSIN-LIKE TYPE B 3.
CC FT DOMAIN 485 518 SAPOSIN-LIKE TYPE A 2.
CC FT DOMAIN 64 139
CC FT DISULFID 67 133 BY SIMILARITY.
CC FT DISULFID 95 107 BY SIMILARITY.
CC FT DISULFID 197 273 BY SIMILARITY.
CC FT DISULFID 200 267 BY SIMILARITY.
CC FT DISULFID 229 240 BY SIMILARITY.
CC FT DISULFID 311 384 BY SIMILARITY.
CC FT DISULFID 314 378 BY SIMILARITY.
CC FT DISULFID 342 353 BY SIMILARITY.
CC FT DISULFID 403 476 BY SIMILARITY.
CC FT DISULFID 406 470 BY SIMILARITY.
CC FT DISULFID 434 445 BY SIMILARITY.
CC FT CARBOHYD 81 81 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 214 214 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 328 328 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 420 420 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CONFLICT 94 94 R -> T (IN REF. 2).
CC FT CONFLICT 486 486 R -> D (IN REF. 2).
CC SEQUENCE 518 AA; 57601 MW; B80300E891C3963 CRC64;

```

Query Match 1.6%; Score 8; DB 1; Length 518;  
 Best Local Similarity 100.0%; Pred. No. 6.1;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 86 VKMETVEK 93  
 |||||  
 Db 293 VKMETVEK 300

```

RESULT 9
ID YFIC_BACSU STANDARD; PRT; 604 AA.
AC P54719;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical ABC transporter ATP-binding protein yf1c.
GN YFIC.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertorello M.G., Bessieres P., Bolotin A., Borcherdt S.,
RA Borriss R., Bourcier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Brueschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dueterhoff A., Ehrlich S.D., Emerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrati E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Giuseppe G., Guy B.J., Haga K., Hatech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holseppel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Kjaer-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Maeda S., Manuel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogihara A., Oudega B., Park S.H.,
RA Paro V., Pohl T.M., Portetle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Ray M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Serot S.J., Serot P., Shin B.S., Soldo B.,
RA Sorokin A., Taccani E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassaretto A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasunoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis."
RT Nature 390:249-256(1997).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MSBA SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D50543; BAA09107.1; -
CC EMBL; Z99108; CAB12651.1; -
CC PIR; P69802; P69802.
CC HSSP; P13569; INBD.
CC Subtilisin; BG1850; YFIC.
CC InterPro; IPR003593; AAA_ATPase.

```

```
DR InterPro: IPR001140; ABC_TM_transpt.  
DR InterPro: IPR003439; ABC_transporter.  
DR Pfam: PF00664; ABC_membrane_1.  
DR Pfam: PF00005; ABC_tran; 1.  
DR ProDom: PD000006; ABC_transporter; 1.  
DR SMART: SM00382; AAA; T.  
DR PROSITE: PS00211; ABC_TRANSPORTER_1; 1.  
DR PROSITE: PS00893; ABC_TRANSPORTER_2; 1.  
KW Hypothetical protein; ATP-binding; Transport; Transmembrane;  
KM Complete proteome.  
FT TRANSMEM 50 70 POTENTIAL.  
FT TRANSMEM 86 106 POTENTIAL.  
FT TRANSMEM 172 192 POTENTIAL.  
FT TRANSMEM 288 308 POTENTIAL.  
FT TRANSMEM 510 530 POTENTIAL.  
FT NP_BIND 399 406 ATP (POTENTIAL).  
SQ SEQUENCE 604 AA; 67396 MW; 1998B4669A2E6133 CRC64;  
  
Query Match 1.4%; Score 8; DB 1; Length 604;  
Best Local Similarity 100.0%; Pred. No. 7;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 40 AAKTANA 47  
DB 474 AAKTANA 481  
|||||  
C555_CHLIT STANDARD; PRT; 86 AA.  
ID C555_CHLIT STANDARD; PRT; 86 AA.  
AC P00133;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE Cytochrome c-555 (C555).  
OS Chlorobium limicola f.sp. thiosulfatophilum.  
OC Bacteria; Chlorobi; Chlorobiales; Chlorobiaceae;  
OC Chlorobium.  
OX NCBI_TaxId=115852;  
RN [1]  
RP SEQUENCE:  
RX MEDLINE=77087088; PubMed=188412;  
RA van Beemmen J., Ambler R.P., Meyer T.E., Olson J.M.,  
RA Shaw E.K.;  
RT "The amino acid sequences of the cytochromes c-555 from two green  
RT sulphur bacteria of the genus Chlorobium.";  
RL Biochem. J. 159:757-774(1976).  
RN [2]  
RP X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS).  
RX MEDLINE=78094383; PubMed=202947;  
RA Korszun Z.R., Saleme F.R.;  
RT "Structure of cytochrome c555 of Chlorobium thiosulfatophilum:  
RT primitive low-potential cytochrome c.";  
RL Proc. Natl. Acad. Sci. U.S.A. 74:5244-5247(1977).  
CC -I- FUNCTION: This basic c-type monoheme cytochrome has been found  
CC exclusively in the green photosynthetic bacteria, although its  
CC role in bacterial photosynthesis is not established. It has an  
CC unusually low redox potential compared with mitochondrial  
CC cytochrome c. It is reactive with cytochrome c oxidases but not  
CC with reductases.  
DR HSP; P11732; 1C5.  
DR InterPro: IPR003088; Cyt_C1.  
DR InterPro: IPR000345; CytC_heme_bind.  
DR Pfam: PF00034; cytochrome_c; 1.  
DR PROSITE: PS00190; CYTOCHROME_C; 1.  
KW Electron transport; Photosynthesis; Heme.  
FT ELECTRON_TRANSPORT; CYTOCHROME_C; 1.  
FT BINDING 14 14 HEME (COVALENT).  
FT BINDING 17 17 HEME (COVALENT).  
FT METAL 18 18 IRON (HEME AXIAL LIGAND).  
FT METAL 60 60 IRON (HEME AXIAL LIGAND).  
SQ SEQUENCE 86 AA; 8780 MW; 0882D21350DB9D4E CRC64;  
  
Query Match 1.4%; Score 7; DB 1; Length 86;
```

```
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 397 AGKATYD 403  
DB 5 AGKATYD 11  
|||||  
C555_CHLITE STANDARD; PRT; 108 AA.  
ID C555_CHLITE STANDARD; PRT; 108 AA.  
AC Q8KG53;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Cytochrome c-555 precursor (C555).  
GN CT0075.  
OS Chlorobium tepidum.  
OC Bacteria; Chlorobi; Chlorobiales; Chlorobiaceae;  
OC Chlorobium.  
OX NCBI_TaxId=1097;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=7LS / ATCC 49652 / DSM 12025;  
RX MEDLINE=22103685; PubMed=12093901;  
RA Eisen J.A., Nelson K.E., Paulsen I.T., Heidelberg J.F., Wu M.,  
RA Dodson R.J., Deboy R., Gwin M.L., Nelson W.C., Haft D.H.,  
RA Hickey E.K., Peterson J.D., Durkin A.S., Kolonay J.L., Yang F.,  
RA Holt I., Umayam L.A., Mason T., Brenner M., Shea T.P., Parksey D.,  
RA Nierman W.C., Feldblyum T.V., Hansen C.L., Craven M.B., Raddue D.,  
RA Vamathevan J.C., Tettelin H., Bryant D.A., Fraser C.M.;  
RT "The complete genome sequence of Chlorobium tepidum 7LS, a  
RT photosynthetic, anaerobic, green-sulfur bacterium.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:9509-9514(2002).  
CC -I- FUNCTION: This basic c-type monoheme cytochrome has been found  
CC exclusively in the green photosynthetic bacteria, although its  
CC role in bacterial photosynthesis is not established. It has an  
CC unusually low redox potential compared with mitochondrial  
CC cytochrome c. It is reactive with cytochrome c oxidases but not  
CC with reductases.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----  
DR EMBL: AE012787; AAM71323.1; -  
DR TIGR: CT0075; -  
DR InterPro: IPR003088; Cyt_C1.  
DR InterPro: IPR002323; Cyt_C1E.  
DR InterPro: IPR000345; CytC_heme_bind.  
DR Pfam: PF00034; cytochrome_c; 1.  
DR PRINTS: PR00607; CYTOCHROME_C.  
DR PROSITE: PS00190; CYTOCHROME_C; 1.  
KW Electron transport; Photosynthesis; Heme; Signal; Complete proteome.  
FT SIGNAL 1 22  
FT CHAIN 23 108  
FT BINDING 36 36 HEME (COVALENT) (BY SIMILARITY).  
FT BINDING 39 39 HEME (COVALENT) (BY SIMILARITY).  
FT METAL 40 40 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).  
FT METAL 82 82 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).  
SQ SEQUENCE 108 AA; 10892 MW; 9711CB0A43B8825C CRC64;  
  
Query Match 1.4%; Score 7; DB 1; Length 108;  
Best Local Similarity 100.0%; Pred. No. 17;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 397 AGKATYD 403  
|||||
```

Db 27 AGKATYD 33

```
RESULT 12
RBS_PROHO
ID_RBS_PROHO STANDARD; PRT; 109 AA.
AC P27569;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Ribulose biphosphate carboxylase small chain (EC 4.1.1.39) (RUBISCO
DE small subunit).
GN CBS OR RBSC.
OS Prochlorothrix hollandica.
OC Bacteria; Cyanobacteria; Prochlorophytes; Prochlorothrixaceae;
OC Prochlorothrix.
OC NCBI_TaxID=1223;
RN [1]
RX MEDLINE=91251137; PubMed=1904095;
RA Morden C.W., Golden S.S.;
RT "Sequence analysis and phylogenetic reconstruction of the gene
RT encoding the large and small subunits of ribulose-1,5-bisphosphate
RT carboxylase/oxygenase from the chlorophyll b-containing prokaryote
RT Prochlorothrix hollandica."
RL J. Mol. Evol. 33:379-395(1991).
CC -1- FUNCTION: RUBISCO catalyzes two reactions: the carboxylation of D-
CC ribulose 1,5-bisphosphate, the primary event in photosynthetic
CC carbon dioxide fixation, as well as the oxidative fragmentation of
CC the pentose substrate in the photorespiration process. Both
CC reactions occur simultaneously and in competition at the same
CC active site.
CC -1- CATALYTIC ACTIVITY: D-ribulose 1,5-bisphosphate + CO(2) = 2 3-
CC phospho-D-glycerate.
CC -1- CATALYTIC ACTIVITY: D-ribulose 1,5-bisphosphate + O(2) =
CC 3-phospho-D-glycerate + 2-phosphoglycolate.
CC -1- SUBUNIT: 8 large chains + 8 small chains.
CC -1- SIMILARITY: BELONGS TO THE RUBISCO SMALL CHAIN FAMILY.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; X57359; CAA0633.1; -
CC DR PIR; S16437;
CC DR InterPro; IPR000894; RUBISCO_small.
CC DR Pfam; PF00101; RUBISCO_small.1.
CC DR PRINTS; PR00152; RUBISCO_SMALL.
CC DR ProDom; PD000290; RUBISCO_small.1.
CC DR Photosynthesis; Carbon dioxide fixation; Photorespiration;
CC Lysase; Oxidoreductase; Monooxygenase.
CC KW
CC SEQUENCE 109 AA; 12963 MW; 9A17AE63ACDD5BE CRC64;
SQ
Query Match 1.4%; Score 7; DB 1; Length 109;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 19 LSDQOIA 25
OY 465 LSDQOIA 471
ID 19 LSDQOIA 25
AC 19 LSDQOIA 25
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
```

```
DE Ribulose biphosphate carboxylase small chain (EC 4.1.1.39) (RUBISCO
DE small subunit).
GN CBS OR RBSC.
OS Synechococcus sp. (strain PCC 7002) (Agmenellum quadruplicatum).
OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
OC NCBI_TaxID=32049;
RN [1]
RX MEDLINE=98196666; PubMed=9537320;
RA Akiyama H., Kanai S., Hirano M., Sugimoto M., Kiyohara M.;
RA Submitted (JUN-1994) to the EMBL/GenBank/DBJ databases.
RL
CC -1- FUNCTION: RUBISCO catalyzes two reactions: the carboxylation of D-
CC ribulose 1,5-bisphosphate, the primary event in photosynthetic
CC carbon dioxide fixation, as well as the oxidative fragmentation of
CC the pentose substrate in the photorespiration process. Both
CC reactions occur simultaneously and in competition at the same
CC active site.
CC -1- CATALYTIC ACTIVITY: D-ribulose 1,5-bisphosphate + CO(2) = 2 3-
CC phospho-D-glycerate.
CC -1- CATALYTIC ACTIVITY: D-ribulose 1,5-bisphosphate + O(2) =
CC 3-phospho-D-glycerate + 2-phosphoglycolate.
CC -1- SUBUNIT: 8 large chains + 8 small chains.
CC -1- SIMILARITY: BELONGS TO THE RUBISCO SMALL CHAIN FAMILY.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; D13971; BAA03078.1; -
CC DR InterPro; IPR000894; RUBISCO_small.
CC DR Pfam; PF00101; RUBISCO_small.1.
CC DR PRINTS; PR00152; RUBISCO_SMALL.
CC DR ProDom; PD000290; RUBISCO_small.1.
CC DR Photosynthesis; Carbon dioxide fixation; Photorespiration;
CC Lysase; Oxidoreductase; Monooxygenase.
CC KW
CC SEQUENCE 111 AA; 13212 MW; 182FA0950AC8EF96 CRC64;
SQ
Query Match 1.4%; Score 7; DB 1; Length 111;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 19 LSDQOIA 25
OY 465 LSDQOIA 471
ID 19 LSDQOIA 25
AC 19 LSDQOIA 25
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein AQ_1239.
GN AQ_1239.
OS Aquifex aeolicus.
OC Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.
OC NCBI_TaxID=63363;
RN [1]
RX MEDLINE=98196666; PubMed=9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Shead M.A., Keller M., Aulay M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus."
RL Nature 352:353-358(1998).
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
```



CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sb-sib.ch](mailto:license@sb-sib.ch)).  
-----  
DR EMBL, AE000730; AAC07249.1; -.  
DR PIR, A70407; A70407.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 111 AA; 13020 MW; 262042D32BD0A3A2 CRC64;  
  
Query Match 1.4%; Score 7; DB 1; Length 111;  
Best Local Similarity 100.0%; Pred. No. 17;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
OY 228 OGLQPPD 234  
Db 28 OGLQPPD 34  
  
RESULT 15  
RS9\_LACIA STANDARD; PRT; 130 AA.  
AC O9CDB7;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE 30S ribosomal protein S9.  
GN RPS1 OR L12253.  
OS *Streptococcus lactis* (subsp. *lactis*) (*Streptococcus lactis*).  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.  
OX NCBI\_TaxID=1360;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=IL1403;  
RX MEDLINE=21235186; PubMed=11337471;  
RA Bolotin A., Wnucker P., Mager S., Jallion O., Malarne K.,  
RA Weisenbach J., Enlrich S.D., Sorokin A.;  
RT "The complete genome sequence of the lactic acid bacterium *Lactococcus*  
RT *lactis* ssp. *lactis* IL1403.";  
RL Genome Res. 11:731-753(2001).  
CC -1- SIMILARITY: BELONGS TO THE S9P FAMILY OF RIBOSOMAL PROTEINS.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sb-sib.ch](mailto:license@sb-sib.ch)).  
-----  
DR EMBL, AE006454; AAK06351.1; -.  
DR PIR, B86906; B86906.  
DR HAMAP, MF\_00532; -; 1.  
DR InterPro; IPR000754; Ribosomal\_S9.  
DR Pfam; PF00380; Ribosomal\_S9; 1.  
DR ProDom; PD001627; Ribosomal\_S9; 1.  
DR PROSITE; PS00360; RIBOSOMAL\_S9; 1.  
KW Ribosomal protein; Complete proteome.  
SQ SEQUENCE 130 AA; 14098 MW; CAF9BC8DCD531BBB CRC64;  
  
Query Match 1.4%; Score 7; DB 1; Length 130;  
Best Local Similarity 100.0%; Pred. No. 20;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
OY 447 GKITYNG 453  
Db 26 GKITYNG 32  
  
RESULT 16  
GREX\_BRUME  
GREX\_BRUME

ID GREX\_BRUME STANDARD; PRT; 157 AA.  
AC O8YID6;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 15-FEB-2003 (Rel. 42, Last annotation update)  
DE Transcription elongation factor greA (Transcript cleavage factor  
DE greA).  
GN GREX OR BME10508 OR BR1504.  
OS Brucella melitensis, and  
OS Brucella suis.  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
OC Brucellaceae; Brucella.  
OX NCBI\_TaxID=29459, 29461;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC SPECIES=B.melitensis; STRAIN=16M / ATCC 23456 / Biotype 1;  
RX MEDLINE=20020109; PubMed=11756688;  
RA DelVecchio V.G., Kapral V., Redkar R.J., Patra G., Mijer C., Los T.,  
RA Ivanova N., Anderson I., Bhattacharya A., Lykidis A., Reznik G.,  
RA Jablonski L., Larsen N., D'Souza M., Bernal A., Mazur M., Goldsman E.,  
RA Selkov E., Elzer P.H., Hagius S., O'Callaghan D., Letesson J.-J.,  
RA Haselkorn R., Kyriides N., Overbeek R.;  
RT "The genome sequence of the facultative intracellular pathogen  
RT *Brucella melitensis*.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC SPECIES=B.suis; STRAIN=1330 / Biovar 1;  
RX MEDLINE=22247741; PubMed=12271122;  
RA Paulsen I.T., Seshadri R., Nelson K.E., Eisen J.A., Heidelberg J.F.,  
RA Read T.D., Dodson R.J., Umayam L., Brinkac L.M., Beanan M.J.,  
RA Daugherty S.C., Deboy R.T., Durkin A.S., Kolonay J.F., Madupu R.,  
RA Nelson W.C., Ayodeji B., Kraul M., Shetty J., Malek J., Van Aken S.E.,  
RA Riedmuller S., Tettelin H., Gill S.R., White O., Salzberg S.L.,  
RA Hoover D.L., Lindler L.E., Halling S.M., Boyle S.M., Frazer C.M.;  
RT "The *Brucella suis* genome reveals fundamental similarities between  
RT animal and plant pathogens and symbionts.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:13148-13153(2002).  
CC -1- FUNCTION: Necessary for efficient RNA polymerase transcription  
CC elongation past template-encoded arresting sites. The arresting  
CC sites in DNA have the property of trapping a certain fraction of  
CC elongating RNA polymerases that pass through, resulting in locked  
CC ternary complexes. Cleavage of the nascent transcript by cleavage  
CC factors such as greA or greB allows the resumption of elongation  
CC from the new 3' terminus. GreA releases sequences of 2 to 3  
CC nucleotides (by similarity).  
CC  
CC -1- SIMILARITY: BELONGS TO THE GREX/GREB FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sb-sib.ch](mailto:license@sb-sib.ch)).  
-----  
DR EMBL, AE009493; AAL51689.1; -.  
DR PIR, AE01445; AAN30415.1; -.  
DR PIR, AF3315; AF3315.  
DR TIGR, BR1504; -.  
DR HAMAP, MF\_00105; -; 1.  
DR InterPro; IPR006359; GreA.  
DR ProDom; PD004918; GreA\_Greb\_N; 1.  
DR Pfam; PF03449; GreA\_Greb\_N; 1.  
DR ProDom; PD004918; GreA\_Greb\_N; 1.  
DR TIGR/Pfam; TIGR01462; greA; 1.  
DR PROSITE; PS00829; GREAB\_1; 1.  
DR PROSITE; PS00830; GREAB\_2; 1.  
KW Transcription regulation; DNA-binding; Coiled coil; Complete proteome.  
FT DOMAIN 46 74 COILED COIL (POTENTIAL).  
SQ SEQUENCE 157 AA; 17528 MW; 5651BAFED6942B3F CRC64;

Query Match 1.4%; Score 7; DB 1; Length 157;  
 Best Local Similarity 100.0%; Pred. No. 24;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 121 EGDTEV 127  
 |||||  
 Db 134 EGDTEV 140

## RESULT 17

KCH\_MOUSE STANDARD; PRT; 172 AA.  
 ID KCH\_MOUSE  
 AC Q920E3;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Voltage-gated potassium channel subfamily H member 5 (Echer-a-go-go potassium channel 2) (Eag2) (Fragment).  
 GN KCNH5 OR EAG2.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6; TISSUE=Brain;  
 RA Saganian M.J., Vega-Saenz de Miera E.C., Rudy B.;  
 RT "Cloning of the mouse eag2 potassium channel".  
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: Pore-forming (alpha) subunit of voltage-gated potassium channel. Elicits a non-inactivating outward rectifying current (by similarity). Channel properties may be modulated by cAMP and subunit assembly.  
 CC -1- SUBUNIT: The potassium channel is probably composed of a homo- or heterotetrameric complex of pore-forming alpha subunits that can associate with modulating beta subunits. Heteromultimer with KCNH1/EAG (by similarity).  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -1- SIMILARITY: Belongs to the potassium channel family. H (Eag) subfamily.  
 CC -1- SIMILARITY: Contains 1 PAS (PER-ARNT-SIM) dimerization domain.  
 CC -1- SIMILARITY: Contains 1 PAS-associated C-terminal (PAC) domain.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AF09565; AAI09442.1; -  
 DR InterPro: IPR001610; PAC.  
 DR InterPro: IPR000700; PAS-assoC.  
 DR InterPro: IPR00014; PAS\_domain.  
 DR Pfam: PF00785; PAC; 1.  
 DR SMART: SMO0086; PAC; 1.  
 DR PROSITE: PS50113; PAC; 1.  
 DR PROSITE: PS50112; PAS; PARTIAL.  
 KW Transport; Ion transport; Ionic channel; Voltage-gated channel;  
 KW Potassium channel; Potassium; Potassium transport; Multigene family.  
 FT NON TER 1 1  
 FT DOMAIN <1 >172 CYTOPLASMIC (BY SIMILARITY).  
 FT DOMAIN <1 59 PAS.  
 FT DOMAIN 60 112 PAC.  
 FT NON TER 172 172  
 SQ SEQUENCE 172 AA; 20064 MM; 99B392D12917E67C CRC64;

Query Match 1.4%; Score 7; DB 1; Length 172;  
 Best Local Similarity 100.0%; Pred. No. 25;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 380 OEAPKTP 386

Db 166 OEAPKTP 172  
 |||||

## RESULT 18

EFA3\_MOUSE STANDARD; PRT; 187 AA.  
 ID EFA3\_MOUSE  
 AC O08545; O55217;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE Ephrin-A3 (EPH-related receptor tyrosine kinase ligand 3) (LERK-3) (EHK1 ligand) (EHK1-L) (Fragment).  
 GN EFN3 OR EPLG3 OR LERK3 OR EPL3.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=129;  
 RX MEDLINE=98126446; PubMed=9465306;  
 RA Cerretti D.P., Nelson N.;  
 RT "Characterization of the genes for mouse LERK-3/Ephrin-A3 (Epl3), mouse LERK-4/Ephrin-A4 (Epl4), and human LERK-6/Ephrin-A2 (Eplg6): conservation of intron/exon structure".  
 RL Genomics 47:131-135 (1998).  
 RN [2]  
 RP SEQUENCE OF 17-78 FROM N.A.  
 RX MEDLINE=97060319; PubMed=8903354;  
 RA Flemmiken A.M., Gale N.W., Yancopoulos G.D., Wilkinson D.G.;  
 RT "Distinct and overlapping expression patterns of ligands for Bsh-related receptor tyrosine kinases during mouse embryogenesis".  
 RL Dev. Biol. 179:382-401 (1996).  
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.  
 CC -1- SIMILARITY: BELONGS TO THE EPHRIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: U93885; AAC39961.1; -  
 DR EMBL: U90666; AAB50241.1; -  
 DR WGI: MG1:106644; Etna3.  
 DR InterPro: IPR001799; Ephrin.  
 DR Pfam: PF00812; Ephrin; 1.  
 DR PRINTS: PR01347; EPHRIN.  
 DR PRODOM: PD002533; Ephrin; 1.  
 DR PROSITE: PS01289; EPHRIN; 1.  
 KW Glycoprotein; GPI-anchor.  
 FT NON TER 1 1  
 FT CARBOHYD 24 24 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 41 41 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 49 49 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CONFLICT 19 19 Y -> I (IN REF. 2).  
 FT CONFLICT 33 34 EQ -> DR (IN REF. 2).  
 FT CONFLICT 46 47 RT -> QP (IN REF. 2).  
 FT CONFLICT 78 78 Y -> W (IN REF. 2).  
 SQ SEQUENCE 187 AA; 21171 MM; CCE4915751760743 CRC64;

Query Match 1.4%; Score 7; DB 1; Length 187;  
 Best Local Similarity 100.0%; Pred. No. 27;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 148 GPGGAB 154  
 |||||  
 Db 27 GPGGAB 33

```

RESULT 19
YEAY_ECOLI
ID YEAY_ECOLI STANDARD; PRT; 193 AA.
AC P76255;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE Hypothetical lipoprotein yeast precursor.
GN YEAY OR B1806 OR C2210.
OS Escherichia coli, and
OS Escherichia coli O6.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
NCBI_TaxID=562, 21792;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RT Science 277:1453-1474(1997).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97251358; PubMed=9097040;
RA Itoh T., Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Iseno K.,
RA Kasai H., Kimura S., Kitakawa M., Kitagawa M., Makino K., Miki T.,
RA Mizouchi K., Mori H., Mori T., Motomura K., Nakade S., Nakamura Y.,
RA Nasahouchi K., Nishio Y., Oshima T., Saito N., Sempel G., Seki Y.,
RA Sivaraman S., Tagami H., Takeda J., Takemoto K., Wada C.,
RA Yamamoto Y., Horuchi T.;
RT "A 460-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 40.1-50.0 min region on the linkage map.";
RT DNA Res. 3:379-392(1996).
[3]
RP SEQUENCE FROM N.A.
RC STRAIN=O6:HI / CFT073 / ATCC 700928;
RX MEDLINE=22388234; PubMed=12471157;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Raab D., Buckles E.L., Liu S.-R., Boutin A.C., Hackett J., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.;
RA Mobley H.L.T., Domeneberg M.S., Blattner F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence
RT of uropathogenic Escherichia coli.";
RT Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
[1- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor
(Potential)].
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC
CC EMBL; AE000275; AAC74876.1; -
CC EMBL; D90824; BAA15610.1; -
CC EMBL; D90825; BAA15615.1; -
CC EMBL; AE016761; AAN80669.1; -
CC PIR; F64941; F64941.
CC Ecogen; EGI3511; Yeay.
CC InterPro; IPR004658; SLP.
CC Pfam; PF03843; SLP; 1.
CC TIGRfam; TIGR00752; slp; 1.
CC PROSITE; PS00013; PROKAR LIPOPROTEIN; 1.
CC Hypothetical protein; Lipoprotein; Membrane; Signal;
CC Complete proteome.
CC
CC FT
CC CHAIN 1 22 POTENTIAL.
CC 23 193 HYPOTHETICAL LIPOPROTEIN YEAY.

```

```

FT LIPID 23 23 N-ACYL DIGLYCERIDE (POTENTIAL).
SQ SEQUENCE 193 AA; 20921 MW; 46CFF28A948A00E CRC64;
Query Match
Best Local Similarity 1.4%; Score 7; DB 1; Length 193;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 17 ALMLSGC 23
Db 17 ALMLSGC 23
-----
RESULT 20
RISE_ARCFU
ID RISE_ARCFU STANDARD; PRT; 194 AA.
AC 027965;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
DE 50S ribosomal protein L15e.
GN RPL15E OR AF2319.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC Archaeoglobaceae; Archaeoglobus.
NCBI_TaxID=2234;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kierlavage A.R., Graham D.E., Kyrpides N.C.,
RA Kirschman R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirschner E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus.";
RT Nature 390:364-370(1997).
[1- SIMILARITY: BELONGS TO THE L15E FAMILY OF RIBOSOMAL PROTEINS.
-----
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC
CC EMBL; AE000944; AAB88937.1; -
CC PIR; G69539; G69539.
CC TIGR; AF2319; -.
CC HAMAP; MF_00256; -; 1.
CC InterPro; IPR000439; Ribosomal_L15e.
CC Pfam; PF00827; Ribosomal_L15e; 1.
CC PROSITE; PS01194; RIBOSOMAL_L15E; 1.
CC Ribosomal protein; Complete proteome.
CC
CC SQ
CC SEQUENCE 194 AA; 23167 MW; 4B607B60F132BD23 CRC64;
-----
Query Match
Best Local Similarity 1.4%; Score 7; DB 1; Length 194;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 76 VDRDHPA 82
Db 134 VDRDHPA 140
-----
RESULT 21

```

RPST MYCTU STANDARD: PRT: 212 AA.

ID RPST MYCTU STANDARD: PRT: 212 AA.

AC 050712;

DT 01-NOV-1997 (Rel. 35, Last Created)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Probable RNA polymerase sigma-D factor.

GN SIGD OR RV3414C OR MT3523 OR MTCY78.15.

OS Mycobacteria tuberculosis.

OC Bacteria: Actinobacteriales; Actinobacteriales; Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.

NCBI\_TaxID=1773;

ON [1]

RP SEQUENCE FROM N.A.

RC STRAIN=H37RV;

RX MEDLINE=96295987; PubMed=9634230;

RA Cole S.T., Broesch R., Parkhill J., Garnier T., Churcher C., Harris D., Gordon S.V., Eigmeier K., Gae S., Barry C.E. III, Tekaia F., Badcock K., Basham D., Brown D., Chillingworth T., Connor R., Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S., Hornsby T., Jagsels K., Krogh A., McLean J., Moule S., Murphy L., Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J., Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E., Taylor K., Whitehead S., Barrett B.G.;

RA "Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence."

RT Nature 393:537-544(1998).

RL [2]

RP SEQUENCE FROM N.A.

RC STRAIN=CDC 1551 / Oshkosh;

RA Fleischmann R.D., Allard D., Eisen J.A., Carpenter L., White O., Peterson J., Deboy R., Dodson R., Gwinn M.L., Haft D., Hickey B., Kolony J.F., Nelson W.C., Weiman L.A., Ermolaeva M.D., Salzberg S.L., Delcher A., Utterback T., Weisman J., Khouri H., Gill J., Mikula A., Bhalal W.;

RA "Whole genome comparison of Mycobacterium tuberculosis clinical and laboratory strains."

RT Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.

RL -1- FUNCTION: THE SIGMA FACTOR IS AN INITIATION FACTOR THAT PROMOTES ATTACHMENT OF THE RNA POLYMERASE TO SPECIFIC INITIATION SITES AND THEN IS RELEASED (BY SIMILARITY).

CC -1- SIMILARITY: BELONGS TO THE SIGMA-70 FACTOR FAMILY. ECF SUBFAMILY.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL: Z77165; CAB01009.1; -

DR EMBL: AE007157; AAK47861.1; ALT\_INIT.

DR PIR: C70737; C70737.

DR TIGR: MT3523; -

DR Tuberculat; RV3414C; -

DR InterPro: IPR000838; Sigma70\_ECF.

DR Pfam: PF04542; sigma70\_r2; 1.

DR Pfam: PF04545; sigma70\_r4; 1.

DR PROSITE: PS01063; SIGMA70\_ECF; 1.

KM Transcription regulation; Sigma factor; DNA-directed RNA polymerase; DNA-binding; Complete proteome.

FT DNA BIND 75 88 POLYMERASE CORE BINDING (POTENTIAL).

FT DNA BIND 176 195 H-T-H MOTIF (BY SIMILARITY).

SO SEQUENCE 212 AA; 22919 MW; 124157F6F66B0C3 CRC64;

Query Match 1.4%; Score 7; DB 1; Length 212; Best Local Similarity 100.0%; Pred. No. 31; Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 488 LSADVA 494

DB 71 LSADVA 77

RESULT 22

ID YIDX\_ECOLI STANDARD: PRT: 218 AA.

AC P31461; P76736;

DT 01-JUL-1993 (Rel. 26, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Hypothetical protein yidx.

GN YIDX OR B3636 OR SF3768.

OS Escherichia coli, and

OS Shigella flexneri.

OC Bacteria: Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.

NCBI\_TaxID=562, 623;

ON [1]

RP SEQUENCE FROM N.A.

RC SPECIES=E.coli; STRAIN=K12 / MG1655;

RX MEDLINE=93315143; PubMed=7686882;

RA Burland V.D., Plunkett G. III, Daniels D.L., Blattner F.R.;

RT "Genome sequence and analysis of 136 kilobases of the Escherichia coli genome: organizational symmetry around the origin of replication."

RL Genome 16:1551-1561(1993).

RM [2]

RP REVISIONS.

RC SPECIES=E.coli; STRAIN=K12 / MG1655;

RX MEDLINE=97426617; PubMed=9278503;

RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V., Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F., Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J., Mau B., Shao Y.;

RT "The complete genome sequence of Escherichia coli K-12."

RL Science 277:1453-1474(1997).

RM [3]

RP SEQUENCE FROM N.A.

RC SPECIES=S.flexneri; STRAIN=301 / Serotype 2a;

RX MEDLINE=22272406; PubMed=12384590;

RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H., Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J., Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S., Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y., Yu J.;

RT "Genome sequence of Shigella flexneri 2a: insights into pathogenicity through comparison with genomes of Escherichia coli K12 and O157."

RL Nucleic Acids Res. 30:4432-4441(2002).

RM [4]

RC CONCEPTUAL TRANSLATION.

RA RUD K.E.;

RL Unpublished observations (JAN-2000).

CC -1- CAUTION: Ref.1 and Ref.2 sequences differ from that shown due to a frameshift in position 188.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL: L10328; AAA62047.1; ALT\_FRAME.

DR EMBL: AE000446; AAC76719.1; ALT\_FRAME.

DR EMBL: AE015384; AAN45211.1; -

DR Ecogene; Egi1719; yidx.

KM Hypothetical protein; Transmembrane; Complete proteome.

FT TRANSMEM 11 31 POTENTIAL.

SO SEQUENCE 218 AA; 24169 MW; 73C7000506883D4 CRC64;

Query Match 1.4%; Score 7; DB 1; Length 218; Best Local Similarity 100.0%; Pred. No. 31; Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 ALMLSGC 23  
 DB 18 ALMLSGC 24

## RESULT 23

TPIS\_METBR STANDARD; PRT; 226 AA.  
 ID TPIS\_METBR 074025;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Triosephosphate isomerase (EC 5.3.1.1) (TIM).  
 GN TPISA OR TPI.  
 OS Methanobacterium bryantii.  
 OC Archaea: Euryarchaeota; Methanobacteria; Methanobacteriales;  
 OC Methanobacteriaceae; Methanobacterium.  
 OX NCBI\_TaxID=2161;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=DSM 862;  
 RA Schramm A.;  
 RN Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.  
 RP CHARACTERIZATION.  
 RX MEDLINE=96198612; PubMed=8925906;  
 RA Kohlhoff M., Dahm A., Hensel R.;  
 RT "Tetrameric triosephosphate isomerase from hyperthermophilic  
 RT Archaea.";  
 RL FEBS Lett. 383:245-250 (1996).  
 CC -1- CATALYTIC ACTIVITY: D-glyceraldehyde 3-phosphate = glyceralone  
 CC phosphate.  
 CC -1- PATHWAY: Plays an important role in several metabolic pathways.  
 CC -1- SUBUNIT: Homodimer.  
 CC -1- SIMILARITY: BELONGS TO THE TRIOSEPHOSPHATE ISOMERASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC \* EMBL: Y11302; CAA72160.1; -  
 DR HAMAP; MF\_00147; -; 1.  
 DR InterPro; IPR003009; FMN enzyme.  
 DR InterPro; IPR000652; Triophos\_ismrse.  
 DR Pfam; PF00121; TIM; 1.  
 DR ProDom; PD001005; Triophos\_ismrse; 1.  
 DR Trigram; TRGR00419; tim; 1.  
 DR PROSITE; PS00171; TIM; FALSE\_NEG.  
 KW Isomerase; Glycolysis; Gluconeogenesis; Fatty acid biosynthesis;  
 KW Pentose shunt.  
 FT ACT SITE 97 BY SIMILARITY.  
 SQ SEQUENCE 226 AA; 23907 MW; 3A6B9DB84A54E546 CRC64;  
 Query Match 1.4%; Score 7; DB 1; Length 226;  
 Best Local Similarity 100.0%; Pred. No. 32;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 256 TGENALK 262  
 DB 21 TGENALK 27

RESULT 24  
 YX09 CAEEL STANDARD; PRT; 230 AA.  
 AC Q1115;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE Hypochemical 26.5 kDa protein C03B1.9 in chromosome X.  
 GN C03B1.9.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;  
 OC Rhabditidae; Pelodierinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RA Martin J.;  
 RN Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC \* EMBL: U40952; AA81738.1; -  
 DR PIR; T15381; T15381.  
 DR WormPep; C03B1.9; C03910.  
 KW Hypochemical protein.  
 SQ SEQUENCE 230 AA; 26473 MW; 082B700AB1229688 CRC64;

Query Match 1.4%; Score 7; DB 1; Length 230;  
 Best Local Similarity 100.0%; Pred. No. 33;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 449 ITVNGNQ 455  
 DB 58 ITVNGNQ 64

RESULT 25  
 EF33 HUMAN STANDARD; PRT; 238 AA.  
 ID EF33 HUMAN  
 AC P52757;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Ephrin-A3 precursor (EPH-related receptor tyrosine kinase ligand 3)  
 DE (LEK3-3) (EHK1 ligand) (EHK1-L).  
 GN EFNA3 OR EPRG3 OR LEK3 OR EFL-2.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95140419; PubMed=7838529;  
 RA Kozlosky C.J., Maraskovsky E., McGrew J.T., Vandenbos T.,  
 RA Teepe M., Lyman S.D., Srinivasan S., Fletcher F.A., Gayle R.B. III,  
 RA Cerretti D.P., Beckmann M.P.;  
 RT "Ligands for the receptor tyrosine kinases hek and elk: isolation of  
 RT cDNAs encoding a family of proteins."  
 RL Oncogene 10:299-306 (1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95063919; PubMed=7973638;  
 RA Davis S., Gale N.W., Aldrich T.H., Maissonpierre P.C., Lhotak V.,  
 RA Pawson T., Goldfarb M., Yancopoulos G.D.;  
 RT "Ligands for EPH-related receptor tyrosine kinases that require  
 RT membrane attachment or clustering for activity."  
 RL Science 266:816-819 (1994).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Ductenum;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,

RA Altschul S.F., Zeeberg B., Buelow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Mariani K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stedman M., Soares M.B., Bonaldo M.F., Casavant T.L., Schetz T.E.,  
 RA Brownstein M.J., Ueda T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,  
 RA Bobak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Gilmwood J., Schmutz J., Myers R.M.,  
 RA Butcherfield Y.S.N., Krzywinski M.I., Skalska U., Smallus D.E.,  
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.,  
 RT "Generation and initial analysis of more than 15,000 full-length  
 human and mouse cDNA sequences".  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.  
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN BRAIN, SKELETAL MUSCLE, SPLEEN,  
 CC THYMUS, PROSTATE, TESTIS, OVARY, SMALL INTESTINE, AND PERIPHERAL  
 CC BLOOD LEUKOCYTES.  
 CC -1- SIMILARITY: BELONGS TO THE EPHRIN FAMILY.  
 CC \*  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: U14187; AAC50078.1; -;  
 DR EMBL: L37360; AAC52368.1; -;  
 DR EMBL: BC017722; AAH17722.1; -;  
 DR PIR: I38849; I38849.  
 DR Genew: HGNC:3223; EFNA3.  
 DR MIM: 601381; -;  
 DR GO: GO:0005887; C:Integral to plasma membrane; TAS.  
 DR GO: GO:0005005; F:transmembrane-ephrin receptor activity; TAS.  
 DR GO: GO:0007287; P:cell-cell signaling; TAS.  
 DR InterPro: IPR00199; Ephrin.  
 DR Pfam: PF00812; Ephrin; 1.  
 DR PRINTS: PR01347; EPHRIN.  
 DR ProDom: PD002533; Ephrin; 1.  
 DR PROSITE: PS01299; EPHRIN; 1.  
 KM Glycoprotein; GPI-anchor; Signal.  
 FT SIGNAL 1 22 POTENTIAL.  
 FT CHAIN 23 238 EPHRIN-A3.  
 FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 67 67 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 100 100 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CONFLICT 71 74 MISSING (IN REF. 2).  
 SQ SEQUENCE 238 AA; 26350 MW; 8EFD6A8BEF33FDDA CRC64;  
 Query Match 1.4%; Score 7; DB 1; Length 238;  
 Best Local Similarity 100.0%; Pred. No. 34;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 148 GPGCGAE 154  
 DB 78 GPGCGAE 84  
 RESULT 26  
 YGIP YEAST STANDARD; PRT; 239 AA.  
 AC P5323;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE Hypothetical 27.6 kDa protein in RPL26B-ACB1 intergenic region.  
 GN YG0303C.

OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=5288;  
 RX MEDLINE=97435481; PubMed=9290212;  
 RA Rieger M., Bruckner M., Schaefer M., Mueller-Auer S.;  
 RT "Sequence analysis of 203 kilobases from Saccharomyces cerevisiae  
 chromosome VII".  
 RL Yeast 13:1077-1090(1997).  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
 CC \*  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: Z72821; CA97024.1; -;  
 DR PIR: S64327; S64327.  
 DR SGD: S0003268; CAX4.  
 DR GO: GO:0030176; C:endoplasmic reticulum membrane, intrinsic P. .; IDA.  
 DR GO: GO:0016462; F:pyrophosphatase activity; IDA.  
 DR GO: GO:0006610; P:lipid biosynthesis; IGI.  
 DR GO: GO:0006487; P:N-linked glycosylation; IMP.  
 DR InterPro: IPR000326; PA\_PTPase.  
 DR Pfam: PF01569; PAP2; 1.  
 DR SMART: SM00014; acidppc; 1.  
 KM Hypothetical protein; Transmembrane.  
 FT TRANSMEM 35 55 POTENTIAL.  
 FT TRANSMEM 132 152 POTENTIAL.  
 FT TRANSMEM 165 185 POTENTIAL.  
 SQ SEQUENCE 239 AA; 27649 MW; C8EC49CBFC08F8F8 CRC64;  
 Query Match 1.4%; Score 7; DB 1; Length 239;  
 Best Local Similarity 100.0%; Pred. No. 34;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 251 SVGALTG 257  
 DB 169 SVGALTG 175  
 RESULT 27  
 FMDC METWP STANDARD; PRT; 272 AA.  
 ID FMDC METWP  
 AC Q3112;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Tungsten-containing formylmethanofuran dehydrogenase II subunit C  
 DE (EC 1.2.99.5).  
 GN FMDC.  
 OS Methanococcus maripaludis.  
 OC Archaea; Euryarchaeota; Methanococci; Methanococcales;  
 OC Methanococcaceae; Methanococcus.  
 OX NCBI\_TaxID=39152;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=JF;  
 RA Yu J.-P., Whitman W.B.;  
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: CATALYZES THE REVERSIBLE OXIDATION OF CO(2) AND  
 CC METHANOFURAN (MFR) TO N-FORMYLMETHANOFURAN (CHO-MFR). THIS ENZYME  
 CC IS OXYGEN-LABILE (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: Formylmethanofuran + H(2)O + acceptor = CO(2)  
 CC + methanofuran + reduced acceptor.  
 CC -1- COFACTOR: TUNGSTEN (BY SIMILARITY).  
 CC -1- PATHWAY: Methanogenesis; first step.

-- SUBUNIT: THIS ENZYME IS COMPOSED OF SEVEN SUBUNITS FWDA (65 kDa),  
CC FWDB (53 kDa), FWDC (31 kDa), FWDD (15 kDa), FWDE, FWDF, AND FWDG  
CC (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE FWDC/FWDC FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: AF029842; AAB82591.1; -  
CC DR InterPro: IPR002489; DUF14.  
CC DR Pfam: PF01493; GKGXG; 1.  
CC Oxidoreductase; Tungsten; Methanogenesis; Repeat.  
CC DOMAIN 77 210 7 X 13 AA REPEATS OF [GW]-X-X-[MLP]-X-X-  
CC G-X-[IL]-X-[IV]-X-G.  
CC FT REPEAT 77 89 1.  
CC FT REPEAT 96 108 2.  
CC FT REPEAT 115 127 3.  
CC FT REPEAT 141 153 4.  
CC FT REPEAT 160 172 5.  
CC FT REPEAT 179 191 6.  
CC FT REPEAT 198 210 7.  
CC SQ SEQUENCE 272 AA; 29174 MW; 50912BD8B47A4BF0 CRC64;  
Query Match 1.4%; Score 7; DB 1; Length 272;  
Best Local Similarity 100.0%; Pred. No. 38;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 447 GKTTVNG 453  
DB 102 GKTTVNG 108  
RESULT 28  
SYGA COXBU STANDARD; PRT; 319 AA.  
ID SYGA COXBU  
AC P94616;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Glycyl-tRNA synthetase alpha chain (EC 6.1.1.14) (Glycine--tRNA ligase  
DE alpha chain) (GLYRS).  
GN GLYO OR CBU913.  
OS Coxiella burnetii.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Legionellales;  
OC Coxiellaceae; Coxiella.  
OX NCBI\_Taxid=777;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Nine Mile phase I / RSA 493;  
RA Williams H., Jaeger C.;  
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Nine Mile phase I / RSA 493;  
RA MEDLINE=22608657; PubMed=12704232;  
RA Seshadri R., Paulsen I.T., Eissen J.A., Read T.D., Nelson K.E.,  
RA Nelson W.C., Ward N.L., Tettelin H., Davidson T.M., Beaman M.J.,  
RA DeBoy R.T., Daugherty S.C., Brinkac L.M., Madupu R., Dodson R.J.,  
RA Khouli H.M., Lee K.H., Carty H.A., Scanlan D., Hainzen R.A.,  
RA Thompson H.A., Samuel J.E., Fraser C.M., Heidelberg J.F.;  
RT "Complete genome sequence of the Q-fever pathogen, Coxiella  
RT burnetii.";  
RT Proc. Natl. Acad. Sci. U.S.A. 100:5455-5460(2003).  
CC -1- CATALYTIC ACTIVITY: ATP + glycine + tRNA(Gly) = AMP + diphosphate  
CC + glycyl-tRNA(Gly).  
CC -1- SUBUNIT: Tetramer of two alpha and two beta chains (By  
CC similarity).  
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.

--1- SIMILARITY: Belongs to class-II aminoacyl-tRNA synthetase family.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: Y10435; CAU71456.1; -  
CC DR EMBL: AE016966; AAC091404.1; -  
CC DR TIGR; CB01913; -  
CC DR HAMAP; MF\_00254; -; 1.  
CC DR InterPro: IPR002310; tRNA\_synth\_2e.  
CC DR InterPro: IPR006194; tRNA\_synth\_Gly.  
CC DR Pfam: PF02091; tRNA\_synth\_2e; 1.  
CC DR PRINTS; PR01044; TRNASYNTHGA.  
CC DR ProDom; PD006985; tRNA\_synth\_2e; 1.  
CC DR TIGRFAMs; TIGR00388; glyo; 1.  
CC DR PROSITE; PS50861; AA tRNA LIGASE II GYAB; 1.  
CC KM Aminoacyl-tRNA synthetase; Protein Biosynthesis; Ligase; ATP-binding.  
CC FT DOMAIN 301 305 305 SER. POLY-SER.  
CC SQ SEQUENCE 319 AA; 36397 MW; FF52579A9F0787F8 CRC64;  
Query Match 1.4%; Score 7; DB 1; Length 319;  
Best Local Similarity 100.0%; Pred. No. 44;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 230 LQPFME 236  
DB 28 LQPFME 34  
RESULT 29  
HAM1\_STRP3 STANDARD; PRT; 328 AA.  
ID HAM1\_STRP3  
AC O8K817;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE HAM1 protein homolog.  
GN SPYM3\_0263 OR SP51596.  
OS Streptococcus pyogenes (serotype M3).  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_Taxid=198466;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MGAS315 / Serotype M3;  
RX MEDLINE=22133808; PubMed=12122206;  
RA Beres S.B., Sylva G.L., Barbican K.D., Lei B., Hoff J.S.,  
RA Mammarella N.D., Liu M.-Y., Smoot J.C., Porcella S.F., Parkins L.D.,  
RA Campbell D.S., Smith T.M., McCormick J.K., Leung D.Y.M.,  
RA Schlievert P.M., Musser J.M.;  
RT "Genome sequence of a serotype M3 strain of group A Streptococcus:  
RT phage-encoded toxins, the high-virulence phenotype, and clone  
RT emergence.";  
RT Proc. Natl. Acad. Sci. U.S.A. 99:10078-10083(2002).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SSI-1 / Serotype M3;  
RA Nakagawa I., Kurokawa K., Nakata M., Tomiyasu Y., Yamashita A.,  
RA Yamazaki K., Okahashi N., Kawabata S., Yasunaga T., Hattori M.,  
RA Hayashi H., Hamada S.;  
RT "The genome of invasive Streptococcus pyogenes; a comparative analysis  
RT of S. pyogenes SSI-1, SF370 and MGAS832.";  
RT Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: BELONGS TO THE HAM1 STRP3E FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; AE014141; AAM7870.1; -  
DR EMBL; AP005146; BAC64691.1; -  
DR HAMAP; MF 01405; fused; 1.  
DR InterPro; IPR002637; Hamip\_1like.  
DR Pfam; PF01725; Hamip\_1like; 1.  
DR ProDom; PD004952; Hamip\_1like; 1.  
DR TIGRFAMs; TIGR00042; TIGR00042; 1.  
KM Hydrolyase; Complete proteome.  
FT DOMAIN 1 129 UNKNOWN.  
SQ SEQUENCE 328 AA; 36395 MW; 91A6D54B6A08888 CRC64;

Query Match 1.4%; Score 7; DB 1; Length 328;  
Best Local Similarity 100.0%; Pred. No. 45;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 366 AVYLPFG 372  
DB 105 AVYLPFG 111

RESULT 30  
HAM1\_STRP8 STANDARD; PRT; 328 AA.  
ID HAM1\_STRP8  
AC Q8P2D2;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DE HAM1 protein homolog.  
GN SPY0362.  
OS Streptococcus pyogenes (serotype M18).  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
CC Streptococcus.  
CC NCBI\_TaxID=186103;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MGAS8232 / Serotype M18;  
RX MEDLINE=21927593; PubMed=11917108;  
RA Smoot J.C., Barlian K.D., Van Gompel J.J., Smoot L.M., Chaussee M.S.,  
RA Sylva G.L., Studevant D.E., Ricklefs S.M., Porcella S.F.,  
RA Pakins L.D., Beres S.B., Campbell D.S., Smith T.M., Zhang Q.,  
RA Kapur V., Daly J.A., Vaasy L.G., Musser J.M.;  
RT "Genome sequence and comparative microarray analysis of serotype M18  
RT group A Streptococcus strains associated with acute rheumatic fever  
RT outbreaks";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:4668-4673(2002).  
CC -1- SIMILARITY: BELONGS TO THE HAM1 NTPASE FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

Best Local Similarity 100.0%; Pred. No. 45;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 366 AVYLPFG 372  
DB 105 AVYLPFG 111

RESULT 31  
HAM1\_STRPY STANDARD; PRT; 328 AA.  
ID HAM1\_STRPY  
AC Q9A1B6;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DE HAM1 protein homolog.  
GN SPY0362.  
OS Streptococcus pyogenes.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
CC Streptococcus.  
CC NCBI\_TaxID=1314;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SF370 / ATCC 700294 / Serotype M1;  
RX MEDLINE=21192684; PubMed=11296296;  
RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,  
RA Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,  
RA Qian X., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,  
RA Yvan X., Clifton S.W., Roe B.A., McLaughlin R.;  
RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes";  
RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).  
CC -1- SIMILARITY: BELONGS TO THE HAM1 NTPASE FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

SQ SEQUENCE 328 AA; 36292 MW; BA7A3677BCAFA254F CRC64;

Query Match 1.4%; Score 7; DB 1; Length 328;  
Best Local Similarity 100.0%; Pred. No. 45;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 366 AVYLPFG 372  
DB 105 AVYLPFG 111

RESULT 32  
PYRB\_LACLE STANDARD; PRT; 351 AA.  
ID PYRB\_LACLE  
AC Q60257;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Aspartate carbamoyltransferase (EC 2.1.3.2) (Aspartate  
DE transcarbamylase) (ATCase).  
GN PYRB.  
OS Lactobacillus leichmannii.  
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;



```

OC Lactobacillus.
OX NCBI_TaxID=28039;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 20076;
RX MEDLINE=96285745; PubMed=8725005;
RA Becker J., Brendel M.;
RT "Molecular cloning and characterization of the pyrB gene of
  Lactobacillus leichmannii encoding aspartate transcarbamylase.";
RL Biochimie 78:3-3(1996).
CC -I- CATALYTIC ACTIVITY: Carbamoyl phosphate + L-aspartate = phosphate
  + N-carbamoyl-L-aspartate.
CC -I- PATHWAY: Pyrimidine biosynthesis; second step.
CC -I- SIMILARITY: BELONGS TO THE ATCASES/OTCASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X84282; CAA59021.1; -
CC EMBL: X78999; CAA55634.1; -
CC PIR: T46956; T46956.
CC HSSP: P00479; 3CSU.
CC HAMAP: MF_00001; -1.
CC InterPro: IPR006130; Asp/Om_Cotranf.
CC InterPro: IPR002082; Asp_Catbmltransf.
CC InterPro: IPR006131; OTCase_O.
CC InterPro: IPR006132; OTCase_P.
CC Pfam: PF00185; OTCase; 1.
CC Pfam: PF02729; OTCase; N: 1.
CC PRINTS: PR00100; AOTCASE.
CC TIGRfam: TIGR00670; asp_carb_tr; 1.
CC PROSITE: PS00097; CARBAMOYLTRANSFERRASE. 1.
CC Pyrimidine biosynthesis; Transferrase.
CC SEQUENCE 351 AA; 39857 MW; 8C36C48B1AEBD956 CRC64;

Query Match      1.4%; Score 7; DB 1; Length 351;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 487 QLSADV 493
DB 46 QLSADV 52

RESULT 33
AVRC_PSESG
ID AVRC_PSESG STANDARD; PRT; 352 AA.
AC F13836;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-FEB-1990 (Rel. 13, Last sequence update)
DT 01-FEB-1991 (Rel. 17, Last annotation update)
DE Avirulence C protein.
GN AVRC
OS Pseudomonas syringae (pv. glycinea).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
  Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=318;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=Race 0;
RX MEDLINE=89008107; PubMed=3049552;
RA Tamaki S., Dahlbeck D., Staakawicz B., Keen N.T.;
RT "Characterization and expression of two avirulence genes cloned from
  Pseudomonas syringae pv. glycinea.";
RL J. Bacteriol. 170:4846-4854(1988).
CC -I- DISEASE: THE TERM AVIRULENT DESCRIBES A POTENTIALLY VIRULENT
  PATHOGEN WHICH IS UNABLE TO INDUCE DISEASE SYMPTOMS IN SPECIFIC
  CULTIVARS OF SOYBEAN.

```

```

CC -I- SIMILARITY: 46% IDENTITY TO AVIRULENCE B PROTEIN OF THE SAME
  STRAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M22219; AAA88428.1; -
CC PIR: B43649; B43649.
CC Virulence.
CC SEQUENCE 352 AA; 39147 MW; B922BF6F41A77A97 CRC64;

Query Match      1.4%; Score 7; DB 1; Length 352;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 364 TDAVYLP 370
DB 249 TDAVYLP 255

RESULT 34
METL_PEA
ID METL_PEA STANDARD; PRT; 374 AA.
AC P49613;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE S-adenosylmethionine synthetase 2 (EC 2.5.1.6) (Methionine
  adenosyltransferase 2) (Adomet synthetase 2).
GN SAM52.
OS Pisum sativum (Garden pea).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
  Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
  eurosid 1; Fabales; Fabaceae; Papilionoideae; Viciales; Pisum.
OX NCBI_TaxID=3888;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Alaska;
RA Gomez L., Carrasco P.;
RL Submitted (OCT-1994) to the EMBL/Genbank/DBJ databases.
CC -I- FUNCTION: CATALYZES THE FORMATION OF S-ADENOSYLMETHIONINE FROM
  METHIONINE AND ATP.
CC -I- CATALYTIC ACTIVITY: ATP + L-methionine + H(2)O = phosphate +
  diposphate + S-adenosyl-L-methionine.
CC -I- COFACTOR: Binds 2 divalent ions, such as magnesium or cobalt, and
  1 potassium ion per subunit (by similarity).
CC -I- PATHWAY: Activated methyl cycle.
CC -I- SIMILARITY: BELONGS TO THE ADOMET SYNTHETASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X82077; CAA57581.1; -
CC EMBL: U36681; AAA58773.1; -
CC PIR: S66352; S66352.
CC HSSP: P04384; IMXB.
CC InterPro: IPR002133; S-Adomet_synt.
CC Pfam: PF00438; S-Adomet_synt; 1.
CC Pfam: PF02772; S-Adomet_synt2; 1.
CC Pfam: PF02773; S-Adomet_synt3; 1.
CC TIGRfam: TIGR01034; meek; 1.
CC PROSITE: PS00376; ADOMET_SYNTHETASE_1; 1.
CC PROSITE: PS00377; ADOMET_SYNTHETASE_2; 1.

```

KM Transferase; One-carbon metabolism; Multigene family; ATP-binding;  
 KM Magnesium; Potassium; Metal-binding.  
 FT NP BIND 121 126 ATP (POTENTIAL).  
 FT METAL 19 19 MAGNESIUM (BY SIMILARITY).  
 FT METAL 45 45 POTASSIUM (BY SIMILARITY).  
 FT METAL 273 273 POTASSIUM (BY SIMILARITY).  
 FT METAL 281 281 MAGNESIUM (BY SIMILARITY).  
 FT BINDING 149 149 ATP (POTENTIAL).  
 SO SEQUENCE 374 AA; 40976 MW; 90921A5AE7420882 CRC64;  
 Query Match 1.4%; Score 7; DB 1; Length 374;  
 Best Local Similarity 100.0%; Pred. No. 50;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 439 SIVANGL 445  
 DB 294 SIVANGL 300  
 RESULT 35  
 ID\_NIR\_RHOSH STANDARD; PRT; 374 AA.  
 AC Q53239;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Copper-containing nitrite reductase precursor (EC 1.7.2.1) (Cu-NIR).  
 GN NRK.  
 OS Rhodospirillum rubrum (Rhodospirillum rubrum).  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodospirales;  
 OC Rhodospiraceae; Rhodospiraceae.  
 OX NCBI\_TaxID=1063;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RC STRAIN=2.4.3;  
 RX MEDLINE=91715533; PubMed=9023188;  
 RA Toques I.E., Kwiatkowski A.V., Shi J., Shapleigh J.P.;  
 RT "Characterization and regulation of the gene encoding nitrite  
 reductase in Rhodospirillum rubrum 2.4.3.";  
 RJ J. Bacteriol. 179:1090-1095(1997).  
 CC -1- CATALYTIC ACTIVITY: Nitrite oxidoreductase  
 CC -1- CATALYTIC ACTIVITY: Nitrite oxidoreductase  
 CC -1- COFACTOR: CONTAINS ONE ATOM OF TYPE 1 COPPER AND ONE ATOM OF TYPE  
 CC IF COPPER: FAD. TYPE 1 COPPER IS BOUND WITHIN A SINGLE MONOMER.  
 CC WHILE TYPE 2 COPPER IS HELD BY RESIDUES FROM EACH OF TWO MONOMERS  
 CC OF THE PRIMER. NITRITE IS BOUND TO THE TYPE 2 COPPER SITE.  
 CC PSEUDODAZURIN IS THE PHYSIOLOGICAL ELECTRON DONOR FOR THE CU-NIR IN  
 CC VITRO (BY SIMILARITY).  
 CC -1- SUBUNIT: Homodimer (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Periplasmic (By similarity).  
 CC -1- DOMAIN: THE TYPE 1 COPPER SITE IN NIR PLAYS A CRUCIAL ROLE FOR  
 CC ELECTRON TRANSFER FROM PSEUDODAZURIN TO THE TYPE 1 COPPER SITE  
 CC OF NIR, WHICH COMPRISES THE CATALYTIC CENTER OF NIR FOR THE  
 CC REDUCTION OF NITRITE.  
 CC -1- SIMILARITY: Contains 2 plastocyanin-like domains.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>  
 CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).  
 CC -----  
 CC EMBL: U62291; AAB05767.1; -  
 CC HSSP: P25006; 1NIF.  
 CC InterPro: IPR001117; Cu-oxidase.  
 CC InterPro: IPR001287; CuNO2\_reductase.  
 CC Pfam: PF00394; Cu-oxidase; 2.  
 CC PRINTS: PR00695; CUO2RDRTASE.  
 CC Oxidoreductase; Copper; Metal-binding; Flavoprotein; FAD;  
 KM Nitrate assimilation; Repeat; Periplasmic; Signal.

FT SIGNAL 1 38 POTENTIAL.  
 FT CHAIN 39 374 COPPER-CONTAINING NITRITE REDUCTASE.  
 FT DOMAIN 93 189 PLASTOCYANIN-LIKE 1.  
 FT DOMAIN 254 355 PLASTOCYANIN-LIKE 2.  
 FT METAL 126 126 COPPER (TYPE 1) (BY SIMILARITY).  
 FT METAL 131 131 COPPER (TYPE 2) (BY SIMILARITY).  
 FT METAL 166 166 COPPER (TYPE 1) (BY SIMILARITY).  
 FT METAL 167 167 COPPER (TYPE 2) (BY SIMILARITY).  
 FT METAL 177 177 COPPER (TYPE 1) (BY SIMILARITY).  
 FT METAL 182 182 COPPER (TYPE 2) (BY SIMILARITY).  
 FT METAL 338 338 COPPER (TYPE 2) (BY SIMILARITY).  
 SO SEQUENCE 374 AA; 40308 MW; 3406B5B5E7DD9934 CRC64;  
 Query Match 1.4%; Score 7; DB 1; Length 374;  
 Best Local Similarity 100.0%; Pred. No. 50;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 142 DFHAATG 148  
 DB 129 DFHAATG 135  
 RESULT 36  
 ID\_NIR\_ALCPA STANDARD; PRT; 376 AA.  
 AC P38501;  
 DT 01-OCT-1994 (Rel. 30, Created)  
 DT 01-OCT-1994 (Rel. 30, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Copper-containing nitrite reductase precursor (EC 1.7.2.1) (Cu-NIR).  
 GN NRK OR NIR.  
 OS Rhodospirillum rubrum (Rhodospirillum rubrum).  
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
 OC Alcaligenaceae; Alcaligenes.  
 OX NCBI\_TaxID=511;  
 RN [1]  
 RC SEQUENCE FROM N.A., AND SEQUENCE OF 34-48 AND 192-240.  
 RC STRAIN=S-6;  
 RX MEDLINE=93294530; PubMed=8515232;  
 RA Nishiyama M., Suzuki J., Kukimoto M., Ohnuki T., Horinouchi S.,  
 RA Beppu T.;  
 RT "Cloning and characterization of a nitrite reductase gene from  
 RT Alcaligenes faecalis and its expression in Escherichia coli.";  
 RJ J. Gen. Microbiol. 139:725-733(1993).  
 RN [2]  
 RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS), AND MUTAGENESIS.  
 RC STRAIN=S-6;  
 RX MEDLINE=94227056; PubMed=8172899;  
 RA Kukimoto M., Nishiyama M., Murphy M.E.P., Turley S., Adman E.T.,  
 RA Horinouchi S., Beppu T.;  
 RT "X-ray structure and site-directed mutagenesis of a nitrite reductase  
 RT from Alcaligenes faecalis S-6: roles of two copper atoms in nitrite  
 RT reduction.";  
 RJ Biochemistry 33:5246-5252(1994).  
 RN [3]  
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).  
 RX MEDLINE=96032602; PubMed=7547950;  
 RA Murphy M.E., Turley S., Kukimoto M., Nishiyama M., Horinouchi S.,  
 RA Sasaki H., Tanokura M., Adman E.T.;  
 RT "Structure of Alcaligenes faecalis nitrite reductase and a copper  
 RT site mutant, M50E, that contains zinc.";  
 RJ J. Biol. Chem. 272:28455-28460(1997).  
 CC -1- CATALYTIC ACTIVITY: Nitrite oxidoreductase  
 CC -1- CATALYTIC ACTIVITY: Nitrite oxidoreductase  
 CC -1- COFACTOR: CONTAINS ONE ATOM OF TYPE 1 COPPER AND ONE ATOM OF TYPE  
 CC IF COPPER: FAD. TYPE 1 COPPER IS BOUND WITHIN A SINGLE MONOMER.

CC WHILE TYPE 2 COPPER IS HELD BY RESIDUES FROM EACH OF TWO MONOMERS  
CC OF THE TRIMER. NITRITE IS BOUND TO THE TYPE 2 COPPER SITE.  
CC PSEUDOAURIN IS THE PHYSIOLOGICAL ELECTRON DONOR FOR THE CU-NIR IN  
CC VITRO (BY SIMILARITY).  
CC -1- PATHWAY: Nitrate assimilation (denitrification).  
CC -1- SUBUNIT: Homotrimer.  
CC -1- SUBCELLULAR LOCATION: Periplasmic.  
CC -1- INDUCTION: UNDER ANAEROBIC GROWTH CONDITIONS AND BY NITRITE.  
CC -1- DOMAIN: THE TYPE 1 COPPER SITE IN NIR PLAYS A CRUCIAL ROLE FOR  
CC ELECTRON TRANSFER FROM PSEUDOAURIN TO THE TYPE II COPPER SITE  
CC OF NIR, WHICH COMPRISES THE CATALYTIC CENTER OF NIR FOR THE  
CC REDUCTION OF NITRITE.  
CC -1- SIMILARITY: Contains 2 plastocyanin-like domains.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: D3155; BAA02440.1; -  
CC PIR: I39582; I39582.  
CC PDB: ZAFN; 01-AUG-96.  
CC PDB: INTD; 08-NOV-96.  
CC PDB: 1A08; 25-FEB-98.  
CC PDB: 1A56; 25-FEB-98.  
CC PDB: 1A57; 25-FEB-98.  
CC PDB: 1A58; 25-FEB-98.  
CC PDB: 1E75; 24-AUG-00.  
CC PDB: 1E77; 24-AUG-00.  
CC PDB: 1E78; 24-AUG-00.  
CC InterPro: IPR001117; Cu-oxidase.  
CC InterPro: IPR001287; CuNO2\_reductase.  
CC InterPro: IPR006311; Tat.  
CC Pfam: PF00394; Cu-oxidase; 2.  
CC PRINTS: PRO0695; CUNO2RDTASE.  
CC TIGRFA: TIGR01409; TAT signal seq; 1.  
CC Oxidoreductase; Copper; Metal-binding; Flavoprotein; PAD;  
CC Nitrate assimilation; Repeat; Periplasmic; Signal; 3D-structure;  
CC Pyroliidone carboxylic acid.  
CC KJ SIGNAL 1 33  
CC FT CHAIN 34 376  
CC FT DOMAIN 34 211  
CC FT MOD\_RES 212 376  
CC FT METAL 131 131  
CC FT METAL 136 136  
CC FT METAL 171 171  
CC FT METAL 172 172  
CC FT METAL 181 181  
CC FT METAL 186 186  
CC FT METAL 342 342  
CC FT MUTAGEN 171 171  
CC FT MUTAGEN 186 186  
CC  
CC COPPER-CONTAINING NITRITE REDUCTASE.  
CC PLASTOCYANIN-LIKE 1.  
CC PLASTOCYANIN-LIKE 2.  
CC PYROLIDONE CARBOXYLIC ACID.  
CC COPPER (TYPE 1).  
CC COPPER (TYPE 2).  
CC COPPER (TYPE 2).  
CC COPPER (TYPE 1).  
CC COPPER (TYPE 1).  
CC COPPER (TYPE 1).  
CC COPPER (TYPE 2).  
CC COPPER (TYPE 2).  
CC H->K: LOOSES NITRITE-REDUCING ACTIVITY.  
CC M->E: CONTAINS ONLY A TYPE II COPPER ATOM  
CC AND FAILS TO CATALYZE THE REDUCTION OF  
CC NITRITE.  
CC  
CC HELIX 42 47  
CC STRAND 50 52  
CC STRAND 58 59  
CC STRAND 67 67  
CC STRAND 74 87  
CC STRAND 90 91  
CC STRAND 94 100  
CC STRAND 101 102  
CC STRAND 108 112  
CC STRAND 113 114  
CC STRAND 116 123  
CC STRAND 125 126  
CC STRAND 131 131  
CC STRAND 134 135  
CC STRAND 136 137

FT HELIX 141 147  
FT STRAND 150 150  
FT TURN 152 153  
FT STRAND 154 161  
FT STRAND 166 171  
FT TURN 175 176  
FT HELIX 178 182  
FT TURN 183 185  
FT STRAND 187 193  
FT TURN 195 195  
FT STRAND 196 196  
FT STRAND 198 199  
FT TURN 201 202  
FT STRAND 205 206  
FT STRAND 210 219  
FT STRAND 223 223  
FT TURN 225 226  
FT STRAND 229 229  
FT HELIX 235 246  
FT TURN 247 249  
FT STRAND 253 256  
FT TURN 257 258  
FT STRAND 259 259  
FT TURN 260 263  
FT HELIX 265 267  
FT STRAND 269 272  
FT TURN 273 274  
FT STRAND 276 283  
FT STRAND 288 288  
FT STRAND 290 293  
FT TURN 294 294  
FT TURN 297 301  
FT STRAND 302 303  
FT TURN 305 305  
FT TURN 306 307  
FT STRAND 311 314  
FT STRAND 319 319  
FT TURN 321 322  
FT STRAND 323 330  
FT STRAND 335 341  
FT HELIX 344 348  
FT TURN 349 349  
FT STRAND 353 359  
FT TURN 364 366  
FT STRAND 367 374  
SQ SEQUENCE 376 AA; 40332 MW; 0AD918988301BF29 CRC64;  
Query Match 1.4%; Score 7; DB 1; Length 376;  
Best Local Similarity 100.0%; Pred. No. 50;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 142 DEHAATG 148  
DB 134 DEHAATG 140  
RESULT 37  
NIR\_ACHCY STANDARD; PRT; 378 AA.  
AC P25006;  
DT 01-MAR-1992 (Rel. 21, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Copper-containing nitrite reductase precursor (EC 1.7.2.1) (Cu-NIR).  
GN NIRK.  
OS Achromobacter cycloclastes.  
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
OC Alcaligenaceae; Achromobacter.  
OX NCBI\_Taxid=223;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=IAM 1013;  
RX MEDLINE=96193667; PubMed=8605003;

RA Chen J.-Y., Chang W.-C., Chang T., Chang W.-C., Liu M.-Y., Payne W.J.,  
RA le Gall J.;  
RT "Cloning, characterization, and expression of the nitric oxide-  
RT generating nitrite reductase and of the blue copper protein genes of  
RT *Achromobacter cycloclastes*.";  
RL Biochem. Biophys. Res. Commun. 219:423-428(1996).  
RN [2]  
RP SEQUENCE OF 39-378.  
RC STRAIN=IAM 1013;  
RX MEDLINE=91308101; PubMed=1830217;  
RA Fenderson F.F., Kumar S., Adman E.T., Liu M.-Y., Payne W.J.,  
RA le Gall J.;  
RT "Amino acid sequence of nitrite reductase: a copper protein from  
RT *Achromobacter cycloclastes*.";  
RL Biochemistry 30:7180-7185(1991).  
RN [3]  
RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).  
RX MEDLINE=91320115; PubMed=1862344;  
RA Godden J.W., Turley S., Teller D.C., Adman E.T., Liu M.-Y.,  
RA Payne W.J., le Gall J.;  
RT "The 2.3-A X-ray structure of nitrite reductase from *Achromobacter*  
RT *cycloclastes*.";  
RL Science 253:438-442(1991).  
RN [4]  
RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).  
RX MEDLINE=96070866; PubMed=7499203;  
RA Adman E.T., Godden J.W., Turley S.;  
RT "The structure of copper-nitrite reductase from *Achromobacter*  
RT *cycloclastes* at five pH values, with NO<sub>2</sub>-bound and with type II  
RT copper depleted.";  
RL J. Biol. Chem. 270:27458-27474(1995).  
CC -1- CATALYTIC ACTIVITY: Nitric oxide + H(2)O + ferriocytochrome c =  
CC nitrite + ferrocytochrome c.  
CC -1- COPACITOR: CONTAINS ONE ATOM OF TYPE I COPPER AND ONE ATOM OF TYPE  
CC II COPPER; PAD.  
CC -1- PATHWAY: Nitrate assimilation (denitrification).  
CC -1- SUBUNIT: Homotrimer.  
CC -1- SUBCELLULAR LOCATION: Periplasmic.  
CC -1- DOMAIN: THE TYPE I COPPER SITE IN NIR PLAYS A CRUCIAL ROLE FOR  
CC ELECTRON TRANSFER FROM PSEUDAZURIN TO THE TYPE II COPPER SITE  
CC OF NIR, WHICH COMPRISES THE CATALYTIC CENTER OF NIR FOR THE  
CC REDUCTION OF NITRITE.  
CC -1- SIMILARITY: Contains 2 plastocyanin-like domains.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; Z48635; CAA88564.1; -.  
DR PIR; JC4648; JC4648.  
DR PDB; 2NRD; 07-DEC-95.  
DR PDB; 1NIA; 07-DEC-95.  
DR PDB; 1NIB; 07-DEC-95.  
DR PDB; 1NIC; 07-DEC-95.  
DR PDB; 1NID; 07-DEC-95.  
DR PDB; 1NIE; 07-DEC-95.  
DR PDB; 1NIF; 07-DEC-95.  
DR InterPro; IPR001117; Cu-oxidase.  
DR InterPro; IPR001287; CUNO2\_reductase.  
DR InterPro; IPR006311; Tat.  
DR Pfam; PF00394; Cu-oxidase; 2.  
DR PRINTS; PR00695; CUNO2RD7ASE.  
DR TIGRfams; TIGR01409; Tat\_signal\_seq; 1.  
KW Oxidoreductase; Copper; Metal-binding; Flavoprotein; PAD;  
KW Nitrate assimilation; Repeat; Periplasmic; Signal; 3D-structure.  
FT SIGNAL 1 38  
FT CHAIN 39 378 COPPER-CONTAINING NITRITE REDUCTASE.  
FT DOMAIN 39 213 PLASTOCYANIN-LIKE 1.  
FT DOMAIN 214 378 PLASTOCYANIN-LIKE 2.

FT METAL 133 133 COPPER (TYPE 1).  
FT METAL 138 138 COPPER (TYPE 2).  
FT METAL 173 173 COPPER (TYPE 2).  
FT METAL 174 174 COPPER (TYPE 1).  
FT METAL 183 183 COPPER (TYPE 1).  
FT METAL 188 188 COPPER (TYPE 1).  
FT METAL 344 344 COPPER (TYPE 2).  
FT METAL 344 344  
FT METAL 47 49  
FT STRAND 52 54  
FT STRAND 60 61  
FT STRAND 69 69  
FT STRAND 76 76  
FT STRAND 89 89  
FT STRAND 92 93  
FT STRAND 96 102  
FT STRAND 103 104  
FT STRAND 105 105  
FT STRAND 110 114  
FT STRAND 115 116  
FT STRAND 118 125  
FT STRAND 127 128  
FT STRAND 133 133  
FT STRAND 136 137  
FT STRAND 138 139  
FT STRAND 143 149  
FT STRAND 152 152  
FT STRAND 154 155  
FT STRAND 156 163  
FT STRAND 168 173  
FT STRAND 177 178  
FT STRAND 180 184  
FT STRAND 185 187  
FT STRAND 189 195  
FT STRAND 197 197  
FT STRAND 198 198  
FT STRAND 200 201  
FT STRAND 203 204  
FT STRAND 207 208  
FT STRAND 212 221  
FT STRAND 225 225  
FT STRAND 227 228  
FT STRAND 231 231  
FT STRAND 237 248  
FT STRAND 249 250  
FT STRAND 255 258  
FT STRAND 259 260  
FT STRAND 261 261  
FT STRAND 262 265  
FT STRAND 267 269  
FT STRAND 271 274  
FT STRAND 275 276  
FT STRAND 278 285  
FT STRAND 290 290  
FT STRAND 292 295  
FT STRAND 296 296  
FT STRAND 299 303  
FT STRAND 304 305  
FT STRAND 308 309  
FT STRAND 313 316  
FT STRAND 321 321  
FT STRAND 323 324  
FT STRAND 325 332  
FT STRAND 337 343  
FT STRAND 346 350  
FT STRAND 351 352  
FT STRAND 355 361  
FT STRAND 366 368  
FT STRAND 368 368  
SQ SEQUENCE 378 AA; 40771 MW; A70B52B814090EA5 CRC64;  
Query Match 1.4%; Score 7; DB 1; Length 378;  
Best Local Similarity 100.0%; Pred. No. 50;  
Matches 7; Conservative 0; Mismatches 0; Gaps 0;  
QY 142 DFHAATG 148



SQ SEQUENCE 387 AA; 41477 MW; A8CAFC0BB27BF14F CRC64;

Query Match 1.4%; Score 7; DB 1; Length 387;

Best Local Similarity 100.0%; Pred. No. 51;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 225 YGEGQLQ 231

DB 135 YGEGQLQ 141

QY 439 SIVANGI 445

DB 292 SIVANGI 298

Search completed: August 27, 2003, 18:43:03  
Job time : 46 secs

## RESULT 40

METK CATRO

ID METK CATRO STANDARD; PRT; 393 AA.

AC Q96551;

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DE 28-FEB-2003 (Rel. 41, Last annotation update)

DE S-adenosylmethionine synthetase 1 (EC 2.5.1.6) (methionine

adenosyltransferase 1) (Adomet synthetase 1).

GN SAMS1.

OS Catharanthus roseus (Rosy periwinkle) (Madagascar periwinkle).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

OC Asteridae; Lamiales; Gentianales; Apocynaceae; Rauvolfioidae; Vinceae;

OC Catharanthus.

OX NCBI\_Taxid=4058;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=9718855; PubMed=9037140;

RA Schroeder G., Eichel J., Breinig S., Schroeder J.;

RT "Three differentially expressed S-adenosylmethionine synthetases from

Catharanthus roseus: molecular and functional characterization.";

Plant Mol. Biol. 33:211-222(1997).

CC -!- FUNCTION: CATALYZES THE FORMATION OF S-ADENOSYLMETHIONINE FROM

METHIONINE AND ATP.

CC -!- CATALYTIC ACTIVITY: ATP + L-methionine + H(2)O = phosphate +

diphosphate + S-adenosyl-L-methionine.

CC -!- COFACTOR: Binds 2 divalent ions, such as magnesium or cobalt, and

1 potassium ion per subunit (By similarity).

CC -!- PATHWAY: Activated methyl cycle.

CC -!- SIMILARITY: BELONGS TO THE ADOMET SYNTHETASE FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL: 271271; CA95856.1; -

DR HSP: P04384; IMXB

DR InterPro: IPR002133; S-Adomet\_synt.

DR Pfam: PF00438; S-Adomet\_synt; 1.

DR Pfam: PF02772; S-Adomet\_syntD2; 1.

DR Pfam: PF02773; S-Adomet\_syntD3; 1.

DR TIGRFAMs: TIGR01034; metK; 1.

DR PROSITE: PS00376; ADOMET SYNTHETASE\_1; 1.

DR PROSITE: PS00377; ADOMET SYNTHETASE\_2; 1.

KW Transferase; One-carbon metabolism; Multigene family; ATP-binding;

KW Magnesium; Potassium; Metal-binding.

FT NP BIND 119 124 ATP (POTENTIAL).

FT METAL 17 17 MAGNESIUM (BY SIMILARITY).

FT METAL 43 43 POTASSIUM (BY SIMILARITY).

FT METAL 271 271 POTASSIUM (BY SIMILARITY).

FT METAL 279 279 MAGNESIUM (BY SIMILARITY).

FT BINDING 147 147 ATP (POTENTIAL).

SQ SEQUENCE 393 AA; 43050 MW; 04371F2B5B386F CRC64;

Query Match 1.4%; Score 7; DB 1; Length 393;

Best Local Similarity 100.0%; Pred. No. 52;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 27, 2003, 18:26:57 ; Search time 102 Seconds

(without alignments)  
1270.024 Million cell updates/sec

Title: US-10-088-045-2

Perfect score: 2626  
Sequence: 1 MSKPTLITKTLICALSALML.....NKGGQLSADVDVAKKXKXPN 502

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

SPREMBL\_23:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phage:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp Vertebrate:\*
- 14: sp Unclassified:\*
- 15: sp\_virus:\*
- 16: sp\_bacteriap:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1487.5	56.6	510	16	Q8XPY3 ralstonia s
2	1388	52.9	390	16	Q9JYE1 neisseria m
3	1386	52.8	386	16	Q9JTB8 neisseria m
4	787	30.0	486	2	Q8KX44 hyphomicrob
5	751.5	28.6	361	1	Q9P9H9 halocarcia
6	481	18.3	360	2	Q9AOC7
7	451	17.2	350	2	Q68601
8	442	16.8	455	16	Q8U7R7
9	439.5	16.7	376	16	Q922Z9
10	431	16.4	376	16	Q8YBB2
11	429.5	16.4	364	2	Q31380
12	429	16.3	376	16	Q8FX29
13	390.5	14.9	478	16	Q8FTV6
14	258	9.8	343	16	Q8FSM1
15	253	9.6	179	2	Q9JTN1
16	250.5	9.5	379	17	Q9HOF4

17	248.5	9.5	338	16	Q8YQ92	Q8YQ92 anabaena sp
18	236	9.0	179	2	Q8RLS7	Q8RLS7 mesorhizobi
19	233	8.9	179	2	Q8RLT4	Q8RLT4 ensifer sp.
20	232	8.8	179	2	Q8RLT3	Q8RLT3 ensifer sp.
21	230.5	8.8	180	2	Q9JTN6	Q9JTN6 nitrosomona
22	228.5	8.7	180	2	Q9JTN3	Q9JTN3 nitrosomona
23	228.5	8.7	180	2	Q9JTN4	Q9JTN4 nitrosomona
24	228	8.7	179	2	Q8RLS6	Q8RLS6 pseudomonas
25	226.5	8.6	180	2	Q9JTN2	Q9JTN2 nitrosomona
26	224.5	8.5	172	2	Q8KPM7	Q8KPM7 uncultured
27	223	8.5	179	2	Q8RLT1	Q8RLT1 ochrobactru
28	222	8.5	138	16	Q8K0J8	Q8K0J8 neisseria m
29	222	8.5	163	16	Q9JVA7	Q9JVA7 neisseria m
30	221	8.4	178	2	Q9JTN5	Q9JTN5 nitrosomona
31	219.5	8.4	180	2	Q9JTN5	Q9JTN5 nitrosomona
32	219	8.3	179	2	Q8RLS8	Q8RLS8 ochrobactru
33	219	8.3	180	2	Q8RT15	Q8RT15 azospirillum
34	216.5	8.2	172	2	Q8KPL0	Q8KPL0 uncultured
35	216	8.2	179	2	Q8RLT0	Q8RLT0 ochrobactru
36	215	8.2	179	2	Q8KPM6	Q8KPM6 uncultured
37	213.5	8.1	172	2	Q8KPM5	Q8KPM5 uncultured
38	211.5	8.1	172	2	Q8KPM3	Q8KPM3 uncultured
39	211.5	8.1	172	2	Q8KPM9	Q8KPM9 uncultured
40	211.5	8.1	172	2	Q8KPM9	Q8KPM9 uncultured
41	210.5	8.0	172	2	Q8KPL80	Q8KPL80 uncultured
42	210.5	8.0	172	2	Q8KPL3	Q8KPL3 uncultured
43	210.5	8.0	172	2	Q8KPM8	Q8KPM8 uncultured
44	210.5	8.0	172	2	Q8KPM8	Q8KPM8 uncultured
45	209.5	8.0	172	2	Q8KPM5	Q8KPM5 uncultured

#### ALIGNMENTS

RESULT 1

Q8XPY3 PRELIMINARY; PRT; 510 AA.

AC Q8XPY3;  
 DT 01-MAR-2002 (TREMBLrel. 20, Created)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
 DE Probable major ANAEROBICALLY induced outer membrane transmembrane  
 DE protein (EC 1.7.99.3).  
 DE RSP1503 OR R503038.  
 OS Ralstonia solanacearum (Pseudomonas solanacearum).  
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
 OC Ralstoniaceae; Ralstonia.  
 OX NCBI\_TaxID=305;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GM11000;  
 RX MEDLINE=21681879; PubMed=11823852;  
 RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,  
 RA Arlat M., Billault A., Broctier P., Camus J.C., Catecolico L.,  
 RA Chandler M., Choise N., Claudel-Renard C., Cunac S., Denape N.,  
 RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,  
 RA Sigler P., Thebaud P., Whalen M., Wincker P., Levy M.,  
 RA Weisenbach J., Boucher C.A.;  
 RT "Genome sequence of the plant pathogen Ralstonia solanacearum";  
 RL Nature 415:497-502(2002).  
 DR EMBL: AL646085; CAD18654.1;  
 DR InterPro: IPR001117; Cu-oxidase.  
 DR InterPro: IPR001287; CUNO2\_reductase.  
 DR InterPro: IPR000345; CytC\_heme\_bind.  
 DR InterPro: IPR003088; CytC.  
 DR Pfam: PF000394; Cytochrome c; 1.  
 DR Pfam: PF000394; Cytochrome c; 1.  
 DR PRINTS: PR00695; CUNO2RPTASE.  
 DR PROSITE: PS00190; CYTOCHROME C; 1.  
 KW Oxidoreductase; Plasmid; Complete proteome.  
 SQ SEQUENCE 510 AA; 54600 MW; 8008105DD99459AC CRC64;





Matches 263; Conservative 45; Mismatches 69; Indels 6; Gaps 2;

QY 7 IKTLLICALSALMLSGCSNQADKAQPKSSTVDAAKTAANADNAASOEHOGLPYIDAIV 66  
 DB 1 MKRQALAAIINAMFALACGGEPPAQOTPAASASASAA---QTAAETPAGSLPIDANT 57  
 QY 67 THAPVPPVDRDHPAKVYVKKETVEKWRMLADGVEYQWTFGGVPGQMIVRGDTIE 126  
 DB 58 THAPVPPVDRDHPAKVYVKKETVEKWRMLADGVEYQWTFGGVPGQMIVRGDTIE 117  
 QY 127 VQFSNHPDSKMPHNVDFHAATPGGGAASFTAPGHTSFFSKALOPGLIYVHCVAAPV 186  
 DB 118 VQFSNHPDSKMPHNVDFHAATPGGGAASFTAPGHTSFFSKALOPGLIYVHCVAAPV 177  
 QY 187 MHIANGMVGLILVEKEGLPKVDKEYVWQGDFTYTKYGEQGLQPFDEMEKAIREDAEV 246  
 DB 178 MHIANGMVGLILVEKEGLPKVDKEYVWQGDFTYTKYGEQGLQPFDEMEKAIREDAEV 237  
 QY 247 VFNQSVGALLTGENALKAKVGETVRLPVNGGPNLTSSPHVIGETFDKVFEGGKSENNHI 306  
 DB 238 VFNQSVGALLTGENALKAKVGETVRLPVNGGPNLTSSPHVIGETFDKVFEGGKSENNHI 297  
 QY 307 QTTLLPAGGAATEKVDVPGDYVVDHAFPAFKKALGLIKVGEENHEIYSHKQDA 366  
 DB 298 QTTLLPAGGAATEKVDVPGDYVVDHAFPAFKKALGLIKVGEENHEIYSHKQDA 357  
 QY 367 VYLPEGAPOAIDTQEAQPTPA 389  
 DB 358 AYAGNGAAPA---ASAPASAPA 377

## RESULT 4

QY 08KKH4 PRELIMINARY; PRT; 486 AA.  
 AC 08KKH4:  
 DT 01-OCT-2002 (TREMBLrel. 22, Created)  
 DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
 DE Nitrite reductase precursor.  
 GN NITR.  
 OS Hyphomicrobium denitrificans.  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 OC Hyphomicrobiales; Hyphomicrobium.  
 OX NCBI\_TaxID=53399;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=AJ3151;  
 RA Kataoka K., Fukui A., Kabayashi M., Yamaguchi K., Suzuki S.;  
 RT "Cloning and expression of copper-containing nitrite reductase from  
 RT Hyphomicrobium denitrificans."  
 RL Submitted (DEC-2001) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; AB076606; BAC00912.1; --  
 DR InterPro; IPR001117; Cu-oxidase.  
 DR InterPro; IPR001287; CUNO2\_reductase.  
 DR PRINTS; PRO0695; CUNO2RDTASE.  
 DR PROSITE; PS00079; MULTICOPPER\_OXIDASE1; 1.  
 KW Signal.  
 FT SIGNAL.  
 SQ SEQUENCE 486 AA; 52238 MW; 6984E02EBCF791AF CRC64;

Query Match 30.0%; Score 787; DB 2; Length 486;  
 Best Local Similarity 51.1%; Pred. No. 5.6e-45;  
 Matches 157; Conservative 41; Mismatches 105; Indels 4; Gaps 3;

QY 55 HOGELPVIDAIVTAP-EVPPVDRDHPAKVYVKKETVEKWRMLADGVEYQWTFGGV 113  
 DB 173 NRAENKSSGADITRDPAOLPEIGRQAKTVIDETVEVKGQOLDNTTYTWTNGKVP 232  
 QY 114 GQMTIRREGDTEVQFSNHPDSKMPHNVDFHAATPGGGAASFTAPGHTSFFSKALOP 173  
 DB 233 GPFLKRVGDTVELHLKHKHKSIMVHSYDFHGAATPGGGAALFTQYDPEBEYVITKALIP 292  
 QY 174 GLVYVHCVAAPVGMHIANGMVGLILVEKEGLPKVDKEYVWQGDFTYTKYGEQGLQPF 233

DB 293 GIYVHCATPSPVPHITTMGMVGLILVEKEGLPKVDKEYVWQGDFTYTKYGEQGLQPF 352  
 QY 234 DMEKAIREDAEVFNQSVGALLTGENALKAKVGETVRLPVNGGPNLTSSPHVIGETFDK 293  
 DB 353 DYKELINKPEYFLFNQSVGALLTGENALKAKVGETVRLPVNGGPNLTSSPHVIGETFDK 412  
 QY 294 VFNQSVGALLTGENALKAKVGETVRLPVNGGPNLTSSPHVIGETFDKVFEGGKSENNHI 351  
 DB 413 VYSLGSVSPPLIGVQTVSVPFGGATIVDFRIDRAGRILVDHALSR-LEHGLVGLVAND 471  
 QY 352 GEENHEI 358  
 DB 472 GPKNDISI 478

## RESULT 5

QY 09P9H PRELIMINARY; PRT; 361 AA.  
 AC 09P9H:  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)  
 DE Copper-containing dissimilatory nitrite reductase precursor  
 DE (EC 1.7.99.3).  
 GN NITR.  
 OS Halobacterium marismortui (Halobacterium marismortui).  
 OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriales;  
 OC Halobacteriaceae; Halobacteriales.  
 OX NCBI\_TaxID=2238;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21311739; PubMed=11418554;  
 RA Ichiki H., Tanaka Y., Mochizuki K., Yoshimatsu K., Sakurai T.,  
 RA Fujiwara T.;  
 RT "Purification, characterization, and genetic analysis of Cu-containing  
 RT dissimilatory nitrite reductase from a denitrifying halophilic  
 RT archaeon, Halobacterium marismortui."  
 RL J. Bacteriol. 183:4149-4156 (2001).  
 DR EMBL; AJ278286; CAB93142.1; --  
 DR HSP; P38501; 1A88.  
 DR InterPro; IPR001117; Cu-oxidase.  
 DR InterPro; IPR001287; CUNO2\_reductase.  
 DR Pfam; PF00394; Cu-oxidase; 1.  
 DR PRINTS; PRO0695; CUNO2RDTASE.  
 KW Oxidoreductase; Signal.  
 FT SIGNAL.  
 FT CHAIN 32 361 POTENTIAL.  
 FT CHAIN 32 361 COPPER-CONTAINING DISSIMILATORY NITRITE  
 FT REDUCTASE.  
 SQ SEQUENCE 361 AA; 38803 MW; 12829C586AA3F5D0 CRC64;

Query Match 28.6%; Score 751.5; DB 1; Length 361;  
 Best Local Similarity 42.5%; Pred. No. 9.2e-43;  
 Matches 154; Conservative 55; Mismatches 138; Indels 15; Gaps 5;

QY 4 PTLIKTTLLICAL--SALMLSGCSNQADKAQPKSSTVDAAKTAANADNAASOEHOGLPY 61  
 DB 5 PTLIKTTLLICAL--SALMLSGCSNQADKAQPKSSTVDAAKTAANADNAASOEHOGLPY 55  
 QY 62 IDAIVTAP-EVPPVDRDHPAKVYVKKETVEKWRMLADGVEYQWTFGGVPGQMIVRG 121  
 DB 56 VDRLAADPTALPDIDSEPTVSEMTTKQVAIEBGVYTYTWTPEQDIPGMIRVR 115  
 QY 122 GDTIEVQFSNHPDSKMPHNVDFHAATPGGGAASFTAPGHTSFFSKALOPGLIYVHC 181  
 DB 116 GDTIEVQFSNHPDSKMPHNVDFHAATPGGGAASFTAPGHTSFFSKALOPGLIYVHC 175  
 QY 182 VAPVGMHIANGMVGLILVEKEGLPKVDKEYVWQGDFTYTKYGEQGLQPFDEMEKAI 241  
 DB 176 VPDNDHSSMFGMILVEKEGLPKVDKEYVWQGDFTYTKYGEQGLQPFDEMEKAI 235  
 QY 242 DAETVFNQSVGALLTGE--NALKAKVGETVRLPVNGGPNLTSSPHVIGETFDKVFEGG 299

Db 236 EPTVLMNGEKXATIPDRHSGSPSMOYGETARVYFVTGPNLDSFPHIGSVMDEWQGS 295  
 Qy 300 KG-ENHNIQTLLIPAGGAIITEFKVDVPGDYLVDAIFRAFNKALCILKVEGENHE 357  
 Db 296 IAGPNRNVQTTVPVKSGSCAIAITLHAIVEGPICKLVDAHSRAKATWAIINREGAND 355  
 Qy 358 IY 359  
 Db 356 VF 357

## RESULT 6

09A007 PRELIMINARY; PRT; 360 AA.  
 ID 09A007; 01-JUN-2001 (TReMBLrel. 17, Created)  
 AC 09A007; 01-JUN-2001 (TReMBLrel. 17, Last sequence update)  
 DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)  
 DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)  
 DE Disimilatory nitrite reductase.  
 GN NIRK.  
 OS Alcaligenes sp. STC1.  
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
 OC Alcaligenaceae; Alcaligenes.  
 NCBI\_TaxID=133923;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=STC1;  
 RA Shoun H., Takaya N.,  
 RT "Alcaligenes sp. STC1 nitrite reductase gene."  
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB046603; BAB21510.1; -.  
 DR HSSP; P25006; INTF.  
 DR InterPro; IPR001117; Cu-oxidase.  
 DR InterPro; IPR001287; Cuno2 reductase.  
 DR Pfam; PF00394; Cu-oxidase; 2.  
 DR PRINTS; PR00695; CUNO2RDTASE.  
 DR PROSITE; PS00283; SOYBEAN KUNITZ; 1.  
 SQ SEQUENCE 360 AA; 39072 MW; CAF063CB38430380 CRC64;

Query Match 18.3%; Score 481; DB 2; Length 360;  
 Best Local Similarity 35.6%; Pred. No. 1.7e-24;  
 Matches 139; Conservative 48; Mismatches 123; Indels 80; Gaps 15;

Qy 3 KPTLLITTLICALSALMLSGCSNQADKAAQPKSTVDAAAKTANADNASQEHOGELPY 62  
 Db 5 RPTLLAALIVTLPASL-----AQAQNA-----DQLPRA 33  
 Qy 63 DATVTAPEVPP--PVDRDHPAKVYVMEVTEVKYMRAD--GVEYQFTFGGOVPGQIRV 119  
 Db 34 KVALVAPPOVHPHEOVAKGSPKVEFTMTTEKKMVIDDKGTTLOANTFSGSNPGPTLV 93  
 Qy 120 REGDTIEVQPSNHPDSKMPHNVDFHAATGPGGAEASFTAPGHTSTFSFKALQPLVYV 179  
 Db 94 HEDDYVELTLVNPATNMPHNVDFHAATGALGAKLTNNVPGQATLRFADRSGFVYV 153  
 Qy 180 CAVAPVGM--HIANGMYGLILVEPKGELP-----KIDKEYVMQGDY----TKGKY 225  
 Db 154 C--APEGMVPWHVVGSGSTLMVLPFDGLDPDGKPLRYDRVYTTIGFDDIYIRDEKGY 211  
 Qy 226 GEGGLQPFMEKAIRED-----AEVVFNGSVGALGENLAKAAGEYVLFV 273  
 Db 212 KDY-----KTLAESYGDVTEVYMRKLTPSHIVFNGKVALTGANNLTKVGETVLLI- 262  
 Qy 274 GNGPNLTSFPHYIGEIFDKVHEFGKGEN--HNIQTLLIPAG--GAATTEFKVDVPG 328  
 Db 263 -HQAQRDRTPHLIGGHDVW--WETGKFGNPPQGNLETWFTIRGSSAALATYFR--QPGV 318  
 Qy 329 YVLVDHAIFFRAFNKALGILKVEGENHEI 358  
 Db 319 YLHLNHLLEAFELGAAGHIKVEGKMWDDL 348

## RESULT 7

068601 PRELIMINARY; PRT; 360 AA.  
 ID 068601; 01-AUG-1998 (TReMBLrel. 07, Created)  
 AC 068601; 01-AUG-1998 (TReMBLrel. 07, Last sequence update)  
 DT 01-AUG-1998 (TReMBLrel. 07, Last sequence update)  
 DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)  
 DE Disimilatory copper-containing nitrite reductase.  
 GN NIR.  
 OS Alcaligenes xylosoxydans (achromobacter xylosoxydans).  
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
 OC Alcaligenaceae; Achromobacter.  
 NCBI\_TaxID=515;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NCIMB11015;  
 RA Suzuki E., Horikoshi N., Kohzuma T.,  
 RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GIFU1051;  
 RA Katoaka K., Furusawa H., Yamaguchi K., Suzuki S.,  
 RT "Cloning and Expression of Copper Nitrite Reductase Gene from  
 RT Alcaligenes xylosoxydans GIFU1051."  
 RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF051831; AAC05831.1; -.  
 DR EMBL; AB013078; BAA33678.1; -.  
 DR HSSP; P25006; INTF.  
 DR InterPro; IPR001117; Cu-oxidase.  
 DR InterPro; IPR001287; Cuno2 reductase.  
 DR Pfam; PF00394; Cu-oxidase; 2.  
 DR PRINTS; PR00695; CUNO2RDTASE.  
 KW Signal.  
 SQ SEQUENCE 360 AA; 38939 MW; 3748B5BD3BF44E7 CRC64;

Query Match 17.2%; Score 451; DB 2; Length 360;  
 Best Local Similarity 35.8%; Pred. No. 1.8e-22;  
 Matches 134; Conservative 45; Mismatches 135; Indels 60; Gaps 15;

Qy 41 AAKTANADNASQEHOGELPYDAITVAPEVPP--PVDRDHPAKVYVMEVTEVKYMR 98  
 Db 18 AAGTMAODA-----DKLPHTKVTLVAPPOVHPHEOATKSPKVEFTMTTEKKMVID 71  
 Qy 99 D-GVEYQFTFGGOVPGQIRVREGDTIEVQPSNHPDSKMPHNVDFHAATGPGGAESF 157  
 Db 72 DKGTTLOANTFSGSNPGPTLVHEDDYVQTLVNPATNMPHNVDFHATGALGAKLTN 131  
 Qy 158 TAPGHTSTFSFKALQPLVYVHCAVAPVGM--HIANGMYGLILVEPKGELP-----K 207  
 Db 132 VNPGEQATLRFADRSGFVYH--APEGMVPWHVVGSGSTLMVLPFDGLDPDGKPLH 189  
 Qy 208 VDKEYVMQGDY----TKGKYGEGLP--QPFMEKAIRED--DAEVVFNQSVGALTGEN 259  
 Db 190 YRAVYTTIGFDDIYIPKPGDKYDAITLAESYGDVQVNRVTLTPSHIVNKGVALTGAN 249  
 Qy 260 ALKAKVGETVRLFVNGSGPNLTSFPHYIGEIFDKVHEFGKGEN--HNIQTLLIPAG-- 314  
 Db 250 ALPAKGETVLLI--HQAQRDRTPHLIGGHDVW--WETGKFGNPPQGNLETWFTIRGSSA 306  
 Qy 315 GAATTEFKVDVPGDYLVDAIFRAFNKALGILKVEGENHEIYSHKOTDAVYLEGAP 374  
 Db 307 GALATYFR--QGVAVAYLHNLLEAFELGAAGHIKVEGKMWND----- 346  
 Qy 375 QAITDQAPKTPAP 388  
 Db 347 ---DLMKQIKAPAP 357

## RESULT 8

Q807R7 PRELIMINARY; PRT; 455 AA.  
 ID Q807R7; 01-JUN-2002 (TReMBLrel. 21, Created)  
 AC Q807R7;  
 DT 01-JUN-2002 (TReMBLrel. 21, Created)

DT 01-JUN-2002 (TRENBLREL. 21, last sequence update)  
 DT 01-MAR-2003 (TRENBLREL. 23, last annotation update)  
 DE Nitrite reductase, copper-containing.  
 GN NIKR OR ATU4382 OR AGR L. 970.  
 OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 OC Rhizobiaceae; Rhizobium.  
 NC NCB1\_Taxid=176299;  
 OK NCB1\_Taxid=176299;  
 RN  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21608550; PubMed=11743193;  
 RA Wood D.W., Serubal J.C., Kaul R., Monks D.E., Kitajima J.P.,  
 RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F., Jr., Woo L.,  
 RA Chen Y., Paulsen I.T., Bisen J.A., Karp P.D., Bovee D. Sr.,  
 RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,  
 RA Kuyavlin T., Levy R., Li M.-J., McCelland E., Palmeri A.,  
 RA Raymond C., Rouse G., Saenphimachak C., Wu Z., Romero P., Gordon D.,  
 RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,  
 RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,  
 RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,  
 RA Nestler E.W.;  
 RT "The genome of the natural genetic engineer Agrobacterium tumefaciens  
 C58.";  
 RT Science 294:2317-2323 (2001).  
 RU  
 RU Science 294:2317-2323 (2001).  
 [2]  
 RN  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21608551; PubMed=11743194;  
 RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,  
 RA Quiclo B., Goldman B.S., Cao Y., Askenazi M., Halling M.,  
 RA Hommel K., Gordon J., Vaudin M., Tatchouk O., Epp A., Liu F.,  
 RA Mollan C., Allinger M., Douglhy D., Scott C., Lapps C., Markelz B.,  
 RA Flanagan C., Crowell C., Gursion J., Lomo C., Sear C., Strub G.,  
 RA Cleto C., Slater S.;  
 RT "Genome sequence of the plant pathogen and biotechnology agent  
 Agrobacterium tumefaciens C58.";  
 RT Science 294:2323-2328 (2001).  
 RU  
 RU EMBL, AE009367; AAL45176.1; ALT\_INIT.  
 DR EMBL, AE008247; AAK89058.1; -  
 DR InterPro: IPR001117; Cu-oxidase.  
 DR InterPro: IPR001287; CuNO2\_reductase.  
 DR InterPro: IPR006311; Tat.  
 DR Pfam: PF00394; Cu-oxidase; 2.  
 DR PRINTS: PRO0695; CUNO2RDTASE.  
 DR TIGRfams: TIGR01409; Tat\_signal\_seq; 1.  
 DR Complete proteome.  
 SQ SEQUENCE 455 AA; 49265 MW; 568D12A8792612B CRC64;  
 Query Match 16.8%; Score 442; DB 16; Length 455;  
 Best Local Similarity 33.2%; Pred. No. 1e-21;  
 Matches 122; Conservative 58; Mismatches 131; Indels 56; Gaps 12;  
 QY 21 SGCSQADKAQPKSTYDAAKTANAD-----NAASOEHQGLPVIDALVTHAPVP 73  
 DB 106 AGAAEKKAAAPLTSALTAALPRAKVDLVKPPVHAHQKAGGPKVVEFTLT----- 159  
 QY 74 PVPDRDHPKVVVKMETEKVRLAD-GVEYQFTFGGVPQOMIRVEGDTIEVOFSNH 132  
 DB 160 -----IK-----EQKMLIDDKGEVHAMFNGSVPRPLVNHODDVELTLINP 203  
 QY 133 PDSKPHVVDFAATGPGGAESTFTAPGHSTFSFKALQPLGLYYHCAVAVGM---HI 189  
 DB 204 DTNELQHNIDFSAAGALGGGLITVNPBEKAILRFKATKAGVFVYHC--APPGVPMVH 261  
 QY 190 ANGMVGLIVEPKELPK-----VDKEYVMQDGFYTK-----GKYGGQGLQPFDM 235  
 DB 262 TSGNMGALVMPREBLDGHGKELYVDKYYVGEODFYIIPDENGNFKKYASAGAMADT 321  
 QY 236 EKALRE-DAEYVFGSVALTGENALAKAVGETVRLFVNGGPNLTSSFVHIGEIFDKV 294  
 DB 332 LEVWKLTPSHIVFNGAVGALGHAHQAAVGE--KVLIVHQAQRDRPHLIGHGIDV 379  
 QY 235 HFEQKGENH--NIQTLLIPAGAAITFEKVDVPGDYVLVDHAIFRAFNKALGILKYE 351

DB 380 -WATKFRNPDLDOETFIPEGTAGAATFTEPGIYAAYVNNHLEAFELGAAAHFKVT 438  
 QY 352 GEENHEI 358  
 DB 439 GEMNNTL 445  
 RESULT 9  
 ID 092229 PRELIMINARY; PRT; 376 AA.  
 AC 092229;  
 DT 01-DEC-2001 (TRENBLREL. 19, Created)  
 DT 01-DEC-2001 (TRENBLREL. 19, last sequence update)  
 DT 01-MAR-2003 (TRENBLREL. 23, last annotation update)  
 DE Putative Nitr Cu-nitrite reductase (EC 1.7.99.3).  
 GN NIKR OR RA0681 OR SMA1250.  
 OS Rhizobium meliloti (Sinorhizobium meliloti).  
 OC Plasmid pSyma (megaplasmid 1).  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 OC Rhizobiaceae; Sinorhizobium.  
 NC NCB1\_Taxid=382;  
 OK NCB1\_Taxid=382;  
 RN  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=1021;  
 RX MEDLINE=21396509; PubMed=11481432;  
 RA Barnett M.J., Fisher R.F., Jones T., Komp C., Abola A.P.,  
 RA Bartoy-Hubler F., Bowser L., Capela D., Galibert F., Gouy J.,  
 RA Gurjal M., Hong A., Huizar L., Hyman R.W., Kahn D., Kahn M.L.,  
 RA Kalman S., Keating D.H., Palm C., Peck M.C., Szurccki R., Wells D.H.,  
 RA Yeh K.-C., Davis R.W., Federspiel N.A., Long S.R.;  
 RT "Nucleotide sequence and predicted functions of the entire  
 Sinorhizobium meliloti pSyma megaplasmid.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 98:9883-9888 (2001).  
 DR EMBL: AE007256; AAK5339.1; -  
 DR InterPro: IPR001117; Cu-oxidase.  
 DR InterPro: IPR001287; CuNO2\_reductase.  
 DR InterPro: IPR006311; Tat.  
 DR Pfam: PF00394; Cu-oxidase; 2.  
 DR PRINTS: PRO0695; CUNO2RDTASE.  
 DR TIGRfams: TIGR01409; Tat\_signal\_seq; 1.  
 DR Oxioreductase; Plasmid; Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 376 AA; 40258 MW; 987999273001DC63 CRC64;  
 Query Match 16.7%; Score 439.5; DB 16; Length 376;  
 Best Local Similarity 33.2%; Pred. No. 1.1e-21;  
 Matches 121; Conservative 52; Mismatches 128; Indels 63; Gaps 13;  
 QY 60 PVIDALVTHADE-----VPPVDRDHP-----AKVVVKMETVE- 92  
 DB 24 PLIGVASHAABEAVALKTAHINVASLPRVKVDLVKPPVHAHQKAGGPKVVEFTLTIE 83  
 QY 93 -KVMRLADGVYQFTFGGVPQOMIRVEGDTIEVOFSNHPSKMPHNVDFHATGFG 151  
 DB 84 KKIIVDEGTBLHANTFNGSVPGPLVNHODDVELTLINPDNTLQHNIDFHSYTGALG 143  
 QY 152 GAESAFTAPGHTSTFSFKALQPLGLYYHCAVAVGM---HIANGMVGLIVEPKGLP- 206  
 DB 144 GGALTVNVPDGTIVLRFASASAGVFVYHC--APPGVPMVHNTSGNMGALMPLPRGLIDG 201  
 QY 207 -----KVDKEYVMQDGFY-----TKKXY-----GEQGLQPFDMKALREDAEYVFGS 251  
 DB 202 KGNSTYDKVYVYVGEODFYVPRDANGKFKYVESGEAVADLTLEVWRTL--TPSHIVFNGA 259  
 QY 232 VGALTGENALAKAVGETVRLFVNGGPNLTSSFVHIGEIFDKVHFEQKGENH--NIQT 308  
 DB 260 VGALTGSDALAAVGE--KVLIVHQAQRDRPHLIGHGIDV- WATKFRNAPVDDET 316  
 QY 309 TLIPAGAAITFEKVDVPGDYVLVDHAIFRAFNKALGILKYEGEENHEIYSHKQTDVAV 368  
 DB 317 WPIPEGTGAATFTEPGIYAAYVNNHLEAFELGAAAHFAVTGDMNDMLMT-----SVR 371  
 QY 369 LPEG 372

Db 372 ADPG 375

## RESULT 10

ID 08YB82 PRELIMINARY; PRT; 376 AA.

AC 08YB82; 01-MAR-2002 (TEMBLrel. 20, Created)

DT 01-MAR-2002 (TEMBLrel. 20, Last sequence update)

DE 01-MAR-2003 (TEMBLrel. 23, Last annotation update)

DE Copper-containing nitrite reductase precursor (EC 1.7.99.3).

GN BMEI10988.

OS Brucella melitensis.

OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;

OC Brucellaceae; Brucella.

OX NCBI\_TaxId=29459;

RN (1)

RP SEQUENCE FROM N.A.

RC STRAIN=16M / ATCC 23456 / Biotype 1;

RX MEDLINE=20020109; PubMed=11756688;

RA Delvecchio V.G., Kapral V., Redkar R.J., Patra G., Mijer C., Los T.,

RA Ivanova N., Anderson I., Bhattacharya A., Lykidis A., Reznik G.,

RA Jablonksi L., Larsen N., D'Souza M., Bernal A., Golsman E.,

RA Selkov E., Elzer P.H., Hagius S., O'Callaghan D., Letesson J.-J.,

RA Haselkorn R., Kyriades N., Overbeek R.;

RT "The genome sequence of the facultative intracellular pathogen

RT Brucella melitensis."

RL Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).

DR EMBL; AF009732; AAL54230.1; -.

DR InterPro; IPR001117; Cu-oxidase.

DR InterPro; IPR001287; CUNO2\_reductase.

DR InterPro; IPR006311; Tat.

DR Pfam; PF00394; Cu-oxidase; 2.

DR PRINTS; PR00695; CUNO2RDTASE.

DR TIGR/Pfam; TIGR01409; TAT\_signal\_seq; 1.

DR Oxidoreductase; Complete proteome.

SQ SEQUENCE 376 AA; 40515 MW; 4D5E3F21C105AF9 CRC64;

Query Match 16.4%; Score 431; DB 16; Length 376;

Best Local Similarity 34.8%; Pred. No. 4.3e-21;

Matches 122; Conservative 48; Mismatches 147; Indels 34; Gaps 11;

Db 43 KTNADNAAQOEQELPVDAITVTHAEPVPRVDHRAKVVVKMETVEKWLAD-GV 101

Db 38 RKASAEIALPRQ-KVELVDPRVHA--HTQVABGPRVVOFTVIEKKIVIDDAGT 93

Qy 102 EYQWTFGGQVPGOMIRVREGDTIEVQFSNHPDSKMPHNVDFHAATGPGGAASFTAPG 161

Db 94 EVHAMTFNGTVPRPLMVHDDVLETLINPETNTLLHIDFHAATGALGGGLTEINPG 153

Qy 162 HTSTPSFKALQPLGYVYHCAVAPVGM--HIANGMGLIVERKEGLP-----KYDKE 211

Db 154 EKTVLRFKATKPGVFVYHC--APPGVPMVHVSGMGAVALVLRGHLDDKGNKLTVDKY 211

Qy 212 YYVMOGDFYRK-----GKYGEQGLQPFMEKAIRE-DAEYVVPNGSVGALTGENAKA 263

Db 212 YVVGEDDFYPRDBNGNYKTYEARGDVETVYKMTLTFTYHVFNGAVALGDAKLT 271

Qy 264 KYGETVRLFVNGGPNLTSSFHVIGELFDKVNHEGKGENHNT--QTTLLPAGGAITEP 321

Db 272 KYGEKVLII--HSQANDTRPHLIGHGIDYVMTAGKENTPRPDVTFPIFGAAAAAFY 329

Qy 322 KVDVPGDYVLVDHAIFFAFKNGKALGLKVGEEHNEIYSHKQDVAIVLPBG 372

Db 330 TFRPGIYAVVNHNLIEAFELGAAHPKVGEMNDLMT-----SLAPSG 375

## RESULT 11

ID 031380 PRELIMINARY; PRT; 364 AA.

AC 031380; 01-JAN-1998 (TEMBLrel. 05, Created)

DT 01-JAN-1998 (TEMBLrel. 05, Last sequence update)

SQ SEQUENCE 364 AA; 39356 MW; D6C197968F290714 CRC64;

DT 01-MAR-2002 (TEMBLrel. 20, Last annotation update)

DE Respiratory nitrite reductase.

GN NITRK.

OS Bradyrhizobium japonicum.

OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;

OC Bradyrhizobiaceae; Bradyrhizobium.

OX NCBI\_TaxId=375;

RN (1)

RP SEQUENCE FROM N.A.

RC STRAIN=USD110;

RA Velasco L., Bedmar E.J.;

RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.

DR EMBL; AJ002516; CA05506.1; -.

DR HSSP; P38501; 2AFN.

DR InterPro; IPR001117; Cu-oxidase.

DR InterPro; IPR001287; CUNO2\_reductase.

DR Pfam; PF00394; Cu-oxidase; 2.

DR PRINTS; PR00695; CUNO2RDTASE.

SQ SEQUENCE 364 AA; 39356 MW; D6C197968F290714 CRC64;

Query Match 16.4%; Score 429.5; DB 2; Length 364;

Best Local Similarity 33.4%; Pred. No. 5.2e-21;

Matches 126; Conservative 50; Mismatches 146; Indels 55; Gaps 12;

Db 14 ALSALMLSGCSNQADKAAQPKSSVDAAKTANADNAAQOEQELPVDAITVTHAEPV 73

Db 12 AATLMTLAPALPADAADKLRQKVELVAPRFVHAHQNTK--QG----- 53

Qy 74 PVVDHRAKVVVKMETVE--KVELADGVEYQFTFGGQVPGOMIRVREGDTIEVQFSN 131

Db 54 -----PKIVPEKLTIEKKVVIDEKGTTQAMTFNGSMGPRPLMVHGGDYVERTLVN 105

Qy 132 HPDSKMPHNVDFHAATGPGGAASFTAPAGHSTSPFKALQPLGYVYHCAVAPVGM--M 187

Db 106 PATNTTPHVIDHSATGALGGALTILNDGEQVLRWKATKTCGVYHC--APGPMIRP 163

Qy 188 HIANGMGLIVERKEGLP-----KYDKEYVMQDGYT---KG---KYGEQGLQPF 233

Db 164 HYVSGMNGAVWVLPBGLNDGKHALKYDKVYVVGQDMVYPRDEKGNFKSYDSPEAST 223

Qy 234 DMEKAIRE-DAEYVVPNGSVGALTGENAKAVGETVRLFVNGGPNLTSSFHVIGELFD 292

Db 224 DTEEMKXKILPHEVFNNGVGTALGTGNALTAVGENV--LIHSAQNRDSRPHLIGHGD 281

Qy 293 KYHFEQKGENH--NIQTTLLPAGGAITEPVDVPGDYVLVDHAIFFAFKNGKALGLK 349

Db 282 YV-WETGKFGNAPVEGLETFIRGSGAGAMTKFMOPGIYAVVTHNLIEADLCAITAPK 340

Qy 350 VEGEENHEIYSHKQDVA 366

Db 341 VEGKNNDDLMTQVAKA 357

## RESULT 12

ID 08FX29 PRELIMINARY; PRT; 376 AA.

AC 08FX29; 01-MAR-2003 (TEMBLrel. 23, Created)

DT 01-MAR-2003 (TEMBLrel. 23, Last sequence update)

DE 01-MAR-2003 (TEMBLrel. 23, Last annotation update)

DE Copper-containing nitrite reductase.

GN BRA0260.

OS Brucella suis.

OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;

OC Brucellaceae; Brucella.

OX NCBI\_TaxId=29461;

RN (1)

RP SEQUENCE FROM N.A.

RC STRAIN=1330 / Biovar 1;

RX MEDLINE=22247741; PubMed=12271122;

RA Paulsen I.T., Seshadri R., Nelson K.E., Eisen J.A., Heidelberg J.F.,

RA Read T.D., Dodson R.J., Umayam L., Brinkac L.M., Beanan M.J.,

RA Daugherty S.C., Debroy R.T., Durkin A.S., Kolonay J.F., Madupu R.,

RA Nelson W.C., Ayodeji B., Kraul M., Shetty J., Malek J., Van Aken S.E.,  
 RA Riedmuller S., Tettele H., Gill S.R., White O., Salzberg S.L.,  
 RA Hoover D.L., Lindler L.E., Halling S.M., Boyle S.M., Fraser C.M.,  
 RT "The Brucella suis genome reveals fundamental similarities between  
 RT animal and plant pathogens and symbionts";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:13148-13153 (2002).  
 DR EMBL: AB014526; AAN33462.1; -.  
 KW Complete proteome.  
 SQ SEQUENCE 376 AA; 40499 MW; 4C0CF7A45FAB1BEA CRC64;

Query Match 16.3%; Score 429; DB 16; Length 376;  
 Best Local Similarity 34.5%; Pred. No. 5.9e-21;  
 Matches 121; Conservative 49; Mismatches 147; Indels 34; Gaps 11;

QY 43 KTNADNNAASOEHOGLPVITDAIVTAPEVPPVDDHAPKVVVKKEVEKWRLLAD-GV 101  
 DB 38 RQASAEIADLRQ-KVELVDPPFVTA---HTQVAGSGKVVQVFTVIEKKIVIDAGT 93  
 QY 102 EYQFTFGQVPGQMIKREGEITVQFSNHPDSKPHNVDPHAATGPGGGAASPTAAG 161  
 DB 94 EYHMTFNGTVPGLVAVVHQQDDVLELTLLINPETNTLLHNI DPHATGALGGGLTEINPG 153  
 QY 162 HTSTSFRLQRLGLVYVHCAVAVPGM---HIANGTGLLVPEKGLP-----KVQKE 211  
 DB 154 EKTVALRFKATKPGVFVYHC--APPGVPMHVVSGMNGAVVVPREGLDHKGKNTLYDYCY 211  
 QY 212 YVVMGQDFYTK-----GKYGEOGLQPFDMEKALRE-DAEYVFNQSYGALTGENALKA 263  
 DB 212 YVVGQDFVPPRDEKNGNTYTPAPGAYEDYVAKWMTLTPTTHVENGAAGALTDGALTA 271  
 QY 264 KVGETVRLVFGNGGPNLTSSFVIGEIFDKVHEGEGKGENHNI--QTLTIPGGAITFE 321  
 DB 272 KSGEKLIT--HSQANRDRTPHLIGHGDIYVATGKFTPPVDQGTWTFPGAAAARFY 329  
 QY 332 KYDVGQDVYLVDAIFRAFNKGAALGLKVEGENHEIYHKQTDVAVLPEG 372  
 DB 330 TFRQGIYAVVYVHNLIEAFELGAVHAFKVTGEMNDLMT-----SLAPSG 375

## RESULT 13

ID 08FTV6 PRELIMINARY; PRT; 478 AA.  
 AC 08FTV6;  
 DT 01-MAR-2003 (TREMBlrel. 23, Created)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
 DE Putative major outer membrane protein.  
 GN CE0272.  
 OS Corynebacterium efficiens.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Corynebacteriaceae; Corynebacteriaceae; Corynebacterium.  
 OX NCBI\_TaxID=152794;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=YS-314 / AJ 12310 / DSM 44549 / JCM 11189;  
 RA Kawarabayashi Y., Yamazaki J., Hino Y., Kikuchi H., Nakamura Y.,  
 RA Ikeo K., Suzuki M., Mashima J., Itoh T., Yamagishi A., Nishio Y.,  
 RA Usuda Y., Sugimoto S.,  
 RT "The entire genome sequence of Corynebacterium efficiens YS-314,"  
 RT Submitted (May-2002) to the EMBL/Genbank/DBJ databases.  
 RL EMBL: AP005214; BAC17082.1; -.  
 DR Complete proteome.  
 KW Complete proteome.  
 SQ SEQUENCE 478 AA; 50578 MW; 282BD1715C02D4A CRC64;

Query Match 14.9%; Score 390.5; DB 16; Length 478;  
 Best Local Similarity 31.4%; Pred. No. 3.3e-18;  
 Matches 108; Conservative 47; Mismatches 130; Indels 59; Gaps 13;

QY 31 AOPKSTVDAAKTAANADNAASOEHOGLPVITDAIVTAPEVPPVDDHAPKVVVKMET 90  
 DB 168 AAPAGCTAPAAADVPGV-GARVSDHGGFTVPVLT-----APALGTVHR---YSMDI 216

QY 91 VEKRWLADGVYQFWTGGQVPGQMIKREGEITVQFSNHPDSKPHNVDPHAA-TGP 149  
 DB 217 TEVEVQVAPGHQVWVLEFGQAPGPTLRGLDTRITL--HNKGTHSIDFHAGEVNP 274  
 QY 150 GGAASPTAPRHTSTFESKALQPLGLVYVHCAVAVGMIANGMGLIVER----KEGL 205  
 DB 275 --DTNMAQTPGESLTVEFVANRYGIMWHCTAPMSLHIANMGAVIIDPADSAESL 332  
 QY 206 PKYDEEYVMQDFYTKGKYGEOGLQPFDMEKALREDAEYVFNQSYGALTGEN----- 259  
 DB 333 SEVDEEYLVASEMELG-----ETELGADQARV--NDRDYDLTAPNFPNRY 377  
 QY 260 ---ALKARVETVRLVFGNGGPNLTSSFVIGEIFDKVHEG-----GKGENHNI 306  
 DB 378 DLAPIEHKVGDTVRVWLMNVGPDQSLSPHVGEVDFDVPSSEGRYLIRDACTRGTGS---- 433  
 QY 307 QTLTIPGGAALTEPKVDVPGDYVLVDAIFRAFNKGAALGLIKV 350  
 DB 434 QAVNVSAQGAFIELTFNAPGNVAFVNHQMTDA-BKGHGFPPTV 476

## RESULT 14

ID 08FSW1 PRELIMINARY; PRT; 343 AA.  
 AC 08FSW1;  
 DT 01-MAR-2003 (TREMBlrel. 23, Created)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
 DE Conserved hypothetical protein.  
 GN LA1555.  
 OS Leptospira interrogans.  
 OC Bacteria; Spirochaetes; Spirochaetales; Leptospiraceae; Leptospira.  
 OX NCBI\_TaxID=173;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=56601 / Serogroup Icterohaemorrhagiae / Serovar lai;  
 RA Ren S.,  
 RL Submitted (MAR-2002) to the EMBL/Genbank/DBJ databases.  
 DR EMBL: AB011334; AAN48754.1; -.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 343 AA; 38017 MW; 6D850CA0F12DB9D3 CRC64;

Query Match 9.8%; Score 258; DB 16; Length 343;  
 Best Local Similarity 28.2%; Pred. No. 1.8e-09;  
 Matches 78; Conservative 41; Mismatches 120; Indels 38; Gaps 11;

QY 75 PVDHDPKAVVVKMETVEKWRLLADGVYQFWTGGQVPGQMIKREGEITVQFSNHPD 134  
 DB 95 PILKQH-----ISIVEMPLVANHNTVAKWTFNGLVPGVRAKLGQRMETLRN--D 145  
 QY 135 SKMPHNVDPHAAATGPG-GGAASPTAPRHTSTFESKALQPLGLVYVHCAVAVGMIANGM 193  
 DB 146 SEHPHSIHFGSHDNEDEGWER--VVQSEKTYQTLAPRIGHPHPCVPPPLASHMAKLG 203  
 QY 194 YGLIVPEKEGLPKYDEEYVMQG-DEYTKGKYGEOGLQPFDMEKALREDAEYVFNQSV 252  
 DB 204 YGGLVDPDPRGRPRNHEMLISGMDLEDKK-----NDLPGMNGMA 245  
 QY 253 GALTGENALKAKVETVRLVFGNGGPNLTSSFVIGEIFDKVHEGEGKGENHNIQTLT 311  
 DB 246 G-FYDRVYIKVPVQGVKLYTIANMCEYEPVASFHLAQTFPV--FRTGTLKILPDHTLV 302  
 QY 312 PAGGA--AITPEKVDVPGDYVLVDAIFRAFNKGAALG 346  
 DB 303 TLGQTERVILDEFTLTPKRGYMFHPQTKMA-EKGAMG 338

## RESULT 15

ID 093TNI PRELIMINARY; PRT; 179 AA.  
 AC 093TNI;  
 DT 01-DEC-2001 (TREMBlrel. 19, Created)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)

DT 01-MAR-2002 (Tremblrel. 20, last annotation update)  
 DE Putative dissimilatory nitrite reductase (Fragment).  
 GN NITR.  
 OS Nitrosomonas sp. TA-9211-NH4.  
 OC Bacteria; Proteobacteria; Betaproteobacteria; Nitrosomonadales;  
 CC Nitrosomonadaceae; Nitrosomonas.  
 NCBI\_TaxID=159355;  
 RX  
 RN  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=TA-9211-NH4;  
 RX MEDLINE=2118534; PubMed=11319103;  
 RA Casatiotti K.L., Ward B.B.;  
 RT "dissimilatory nitrite reductase genes from autotrophic ammonia-  
 oxidizing bacteria";  
 RL Appl. Environ. Microbiol. 67:2213-2221 (2001).  
 DR EMBL; AF339049; AK53363.1;  
 DR InterPro; IPR001117; Cu-oxidase.  
 DR Pfam; PF00394; Cu-oxidase; 1.  
 FT NON\_TER 1 179  
 FT SEQUENCE 179 AA; 20005 MW; 85AOCFEF7963B12D CRC64;  
 SQ  
 Query Match 9.6%; Score 253; DB 2; Length 179;  
 Best Local Similarity 39.6%; Pred. No. 1.6e-09;  
 Matches 74; Conservative 23; Mismatches 58; Indels 32; Gaps 11;  
 QY 177 VYHCANAVPM--HANGMYGLIVPEKGLPK-----VDKEYVMQGDY----TK 222  
 DB 1 VYHC-APGGMVPMHVGSMGTIMVLPKGLKPKGLHYDRIVYIGENDYIPKDN 58  
 QY 223 GKY-----GEOGLQFPD-MEKAIREDAYVFNVSAGALTKENAKAKVGETVRLFVGN 275  
 DB 59 GKITYVDSPEAPADTTDIRKLI---PNHVFNKRGVAGLTGNKATKAGETV-MFV-H 113  
 QY 276 GGFNLTSFVIGEIFDKVHFEGSGKEN--HNIOITLLPAGGAITEFKVDVPGDYVLV 332  
 DB 114 SEKNRDTIRPHIIGLDYV-WETGKFSNPEKDLFTFMFIGSAGALYFRQPGYVAVY 172  
 QY 333 DHAIFRA 339  
 DB 173 NHNLIEA 179  
 ID PRELIMINARY; PRT; 379 AA.  
 AC O9HQF4;  
 DT 01-MAR-2001 (Tremblrel. 16, Created)  
 DT 01-MAR-2001 (Tremblrel. 16, last sequence update)  
 DT 01-MAR-2003 (Tremblrel. 23, last annotation update)  
 GN Membrane protein.  
 DE PANI OR VNG1187G.  
 OS Halobacterium sp. (strain NRC-1).  
 OC Archaea; Euryarchaeota; Halobacteriales;  
 CC Halobacteriaceae; Halobacterium.  
 NCBI\_TaxID=64091;  
 RX  
 RN  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=20504483; PubMed=11016950;  
 RX NG W.V., Kennedy S.P., Mahaias G.G., Bergquist B., Pan M.,  
 RA Shukla H.D., Lesky S.R., Baliga N.S., Thorsson V., Shroga J.,  
 RA Swartzell S., Weir D., Hall J., Dahl T.A., Meli R., Goo Y.A.,  
 RA Leitauer B., Keller K., Cruz R., Danson M.J., Hough D.W.,  
 RA Maddocks D.G., Jablonki P.E., Krebs M.P., Angevine C.M., Dale H.,  
 RA Isenbarger T.A., Peck R.F., Pohlischer M., Spudich J.L., Jung K.-H.,  
 RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,  
 RA Ehardt H., Lowe T.M., Liang P., Riley M., Hood L., Dasgupta S.;  
 RT "Genome sequence of Halobacterium species NRC-1";  
 RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181 (2000).  
 DR EMBL; AE005046; AAC19561.1;  
 DR InterPro; IPR001117; Cu-oxidase.  
 DR Pfam; PF00394; Cu-oxidase; 2.

DR TIGRfams; TIGR01409; TAT\_signal\_seq; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 379 AA; 41173 MW; 4807DF07F8AA666 CRC64;  
 Query Match 9.5%; Score 250.5; DB 17; Length 379;  
 Best Local Similarity 25.3%; Pred. No. 6.8e-09;  
 Matches 94; Conservative 52; Mismatches 133; Indels 93; Gaps 20;  
 QY 20 LSGCNSQADKAKAPSSIVDAAK-----TANADNAASQHOGELEVIDAIVTHAREVPP 75  
 DB 29 IAGCT-APDGEVTDITPAKQSGLPPTSPEVVDATGQNOVTLKSV-----PA 78  
 QY 76 VDRDPAKV--VMEYEVKMYRLADVEYQFTFGGQ-----VPGMIRVREDITTEV 127  
 DB 79 VHDVPLDSMGSPVLTPLPV-----WAFATEDGSPVPGPIVRETEGODLE 124  
 QY 128 QFSNHPDSKMPHNVDFHAATGPGGAENASF-----TANGHTSTSFKALQRL 175  
 DB 125 TLDN-TDGKRPHTLHFH-----GSQTAMEDDGVPTTGIRVGEKRTYITIPANVPGT 176  
 QY 176 VYHCANAVPMHANGMYGLIVPEKGLPKVDKEYVMQGDYFTKGYEGQLQFPDM 235  
 DB 177 HLHYCHY-QTORHIDMGMYGIFRIDPK-GYEPADKYEFTVTKWDMSRLKSNAG----- 228  
 QY 236 EKAIREDAYV--VENSAGALTGENA-----LKAKVGETVRLFVNGGPNLTS 282  
 DB 229 -----EDVDYSPRTNPNPVFTVNGKSAPRTLHPEDGSPVIVQDPTVRLHVNNG-YNMH 282  
 QY 283 SFHVIGEIFDKVHFEGK--GENNIO-TLLIPAGGAITEFKVDV-PGDYVL----VD 333  
 DB 283 PLHINHRRFMYEKQGGQIPAAQHTMTDITMAPARHTI-EFOADADPGIYIMHCHKVN 341  
 QY 334 HAIFRAFNKAL 345  
 DB 342 HVNMGTFYPGGM 353  
 ID PRELIMINARY; PRT; 338 AA.  
 AC O8YO92;  
 DT 01-MAR-2002 (Tremblrel. 20, Created)  
 DT 01-MAR-2002 (Tremblrel. 20, last sequence update)  
 DT 01-MAR-2003 (Tremblrel. 23, last annotation update)  
 DE Hypothetical protein A113942.  
 GN ALU3942.  
 OS Anabaena sp. (strain PCC 7120).  
 OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.  
 NCBI\_TaxID=103690;  
 RX  
 RN  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=21595285; PubMed=11759840;  
 RX Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,  
 RA Watanabe A., Iriuguchi M., Ishikawa A., Kawashima K., Kimura T.,  
 RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,  
 RA Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,  
 RA Yasuda M., Tabata S.;  
 RT "Complete genomic sequence of the filamentous nitrogen-fixing  
 cyanobacterium Anabaena sp. strain PCC 7120";  
 RL DNA Res. 8:205-213 (2001).  
 DR EMBL; AP003594; BAB75641.1;  
 DR InterPro; IPR001117; Cu-oxidase.  
 DR InterPro; IPR001287; Cuno2\_reductase.  
 DR Pfam; PF00394; Cu-oxidase; 1.  
 DR PRINTS; PR00695; CUNO2RDTASE.  
 DR Hypothetical protein; Complete proteome.  
 SO SEQUENCE 338 AA; 38408 MW; AFBRI1BB9D01627F7 CRC64;  
 Query Match 9.5%; Score 248.5; DB 16; Length 338;  
 Best Local Similarity 27.8%; Pred. No. 7.8e-09;  
 Matches 76; Conservative 42; Mismatches 116; Indels 39; Gaps 12;  
 QY 94 VMRLADGVYQFWTFGQVPGQMIRVREDTIEVQFSNHPDSKMPHNVDFHAATGPGGA 153

Db 93 VQOLNSAVSNINWDLNGRIPLPGTLTKAKQGRIRIVLF--NHQASHSLHFGV----HBA 146  
Qy 154 EASFTAP---GHTSTSPKALOPGLVYVHCAVAPVGMHTANGVGLIVEPKGGLPKVX 210  
Db 147 EMDGVRPVNNNSMTIYEPDAPRYGVHLHYCHIEPVRHAKGLYGMFTIDPPRPAPDE 206  
Qy 211 EYVWQGDFTYTKGKGEGLOPFDMEKARREDAVYVFGSVALTGENALAKAVGETVA 270  
Db 207 IYLVNMG-----YVDNDSHND--FAFNGLPYHNM-DNPICQYONQLR 248  
Qy 271 LEVGN---GGNLTSSFHYIGELFDKAVHEGGKGENHNIQTLLIPAGAA--ITEPKVDV 325  
Db 249 LYVNLINIEYDPAPT--FHLHAFNFDYRY--GMSMKASEKTIVITGVAERHILEFAPRY 304  
Qy 326 PGDYVLVDHAIFFAFKAGALILKY--EGEENH 356  
Db 305 PGKYMFPHQDALAEN-GCMGQFEVAVANNONH 336

## RESULT 18

Q8RLT3 PRELIMINARY; PRT; 179 AA.  
ID Q8RLT3; 01-JUN-2002 (Tremblrel. 21, Created)  
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)  
DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)  
DE Putative dissimilatory nitrite reductase (Fragment).  
GN NIRK.  
OS Mesorhizobium sp. 4FB1.  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
OC Rhizobiaceae; Mesorhizobium.  
OC NCBI\_TaxID=126151;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=4FB1;  
RA Song B., Ward B.B.;  
RT "Nitrite reductase genes in halobenzate degrading denitrifying  
RT bacteria."  
RL Submitted (FE8-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AY078254; AAL82513.1; -  
DR InterPro; IPR001117; Cu-oxidase.  
DR Pfam; PF00394; Cu-oxidase; 1.  
FT NON\_TER 1  
FT SEQUENCE 179 AA; 19893 MW; 0060433CAD0A516C CRC64;

Query Match 9.0%; Score 236; DB 2; Length 179;  
Best Local Similarity 37.8%; Pred. No. 2.2e-08;  
Matches 70; Conservative 27; Mismatches 60; Indels 28; Gaps 10;

Qy 177 VVHCAVAPVGM---HIANGVGLIVEPKGGLP-----KYDKEYVWQGDFTYTKGKG 226  
Db 1 VVHC--APGGMTPMVVSGMHTIVLPPDGLKNEKGLVYKDYIYIDNFTYIP-RDE 57  
Qy 227 ECGLOPFD-MEKARREDAE-----YVFGSVALTGENALAKAVGETVRLFPVNGG 277  
Db 58 DGNFKKFDLSIGESYEETMEVWGLIPTHVVFNGRGSLTDNAMSKEVEIV-LFY-HSQ 115  
Qy 278 PNLTSFHYIGELFDKAVHEGGKGEN--HNIQTLLIPAGAAITEFKVDVPGDYVLVDH 334  
Db 116 ANRDRPRLHIGHDYV--WEGKFAKPAKDLFTWIFRGSAGALYTLQGVAVVNH 174  
Qy 335 AIFRA 339  
Db 175 NLIEA 179

RESULT 19  
Q8RLT4 PRELIMINARY; PRT; 179 AA.  
ID Q8RLT4; 01-JUN-2002 (Tremblrel. 21, Created)  
DT 01-JUN-2002 (Tremblrel. 21, Created)

DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)  
DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)  
DE Putative dissimilatory nitrite reductase (Fragment).  
GN NIRK.  
OS Ensifer sp. 2FB8.  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
OC Rhizobiaceae; Ensifer.  
OC NCBI\_TaxID=126147;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=2FB8;  
RA Song B., Ward B.B.;  
RT "Nitrite reductase genes in halobenzate degrading denitrifying  
RT bacteria."  
RL Submitted (FE8-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AY078247; AAL82506.1; -  
DR InterPro; IPR001117; Cu-oxidase.  
DR Pfam; PF00394; Cu-oxidase; 1.  
FT NON\_TER 1  
FT SEQUENCE 179 AA; 19695 MW; D94A75486978BD7A CRC64;

Query Match 8.9%; Score 233; DB 2; Length 179;  
Best Local Similarity 38.6%; Pred. No. 3.5e-08;  
Matches 71; Conservative 22; Mismatches 65; Indels 26; Gaps 8;

Qy 177 VVHCAVAPVGM---HIANGVGLIVEPKGGLP-----KYDKEYVWQGDFTYTKG 222  
Db 1 VVHC--APGGMTPMVVSGMHTIVLPPDGLKDEKQPLTYDKIYVGEODFYVPKDEA 58  
Qy 223 ---GKYGEGLOPFDMEKARE-DAEYVFGSVALTGENALAKAVGETVRLFPVNGG 278  
Db 59 GNYKYETPGEAYEDAVVAMRLTPTHVFNAGVALTDHALTAAGE--RVLVHNSA 116  
Qy 279 NLTSSFHYIGELFDKAVHEGGKGENH--NIQTLLIPAGAAITEFKVDVPGDYVLVDH 335  
Db 117 NRDTPRLHIGHDYV-WATGKFRNPDLDETWLIPGTAAGAAFTFPQGVAVVNH 175  
Qy 336 IFRA 339  
Db 176 LIEA 179

RESULT 20  
Q8RLT3 PRELIMINARY; PRT; 179 AA.  
ID Q8RLT3; 01-JUN-2002 (Tremblrel. 21, Created)  
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)  
DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)  
DE Putative dissimilatory nitrite reductase (Fragment).  
GN NIRK.  
OS Ensifer sp. 4FB6.  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
OC Rhizobiaceae; Ensifer.  
OC NCBI\_TaxID=126148;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=4FB6;  
RA Song B., Ward B.B.;  
RT "Nitrite reductase genes in halobenzate degrading denitrifying  
RT bacteria."  
RL Submitted (FE8-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AY078248; AAL82507.1; -  
DR InterPro; IPR001117; Cu-oxidase.  
DR Pfam; PF00394; Cu-oxidase; 1.  
FT NON\_TER 1  
FT SEQUENCE 179 AA; 19594 MW; B9EA63390B6CF6E CRC64;

Query Match 8.8%; Score 232; DB 2; Length 179;  
Best Local Similarity 38.2%; Pred. No. 4.1e-08;  
Matches 71; Conservative 20; Mismatches 65; Indels 30; Gaps 8;

```

QY 177 VYHCAVAPVGM---HIANGMYGLILVEPKEGLP-----KVDKEYYVMQGDIFYK----- 222
DB 1 VYHC--APGGMVPMWHTSGMNGAIWVLPDGLKDEKQGPLTYDKIYVGEOPFYPKDEA 58
QY 223 ---GKYGEQGLQPFDMKEAIR-DAEYVFNKSVGALTGEMNAKAKVGETVRLFVNGG 278
DB 59 GNYKKTETGEAYEDVAKMRILTPHYFNKVGALTDGHALTAAGV--KVLVYHSAQ 116
QY 279 NLTSSEFHVIGEIFDKVHFEKGKGNH-----IQTLLIPAGGAITEFKVDVPGDYVLVD 333
DB 117 NADTRPHLIGGHQDYV---GATGCFKFNPPDLQDETLLIPGTAAGAFYTRQPGVAYAVN 173
QY 334 AIFRA 339
DB 174 NMLIEA 179

RESULT 21
ID Q93TN6 PRELIMINARY; PRT; 180 AA.
AC Q93TN6;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE Putative dissimilatory nitrite reductase (Fragment).
GN Nitr.
OS Nitrosomonas marina.
OC Bacteria; Proteobacteria; Betaproteobacteria; Nitrosomonadales;
OC Nitrosomonadaceae; Nitrosomonas.
OX NCBI_Taxid=917;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-56;
RC MEDLINE=21218534; PubMed=11319103;
RA Casciotti K.L., Ward B.B.;
RT "Dissimilatory nitrite reductase genes from autotrophic ammonia-oxidizing bacteria.";
RL Appl. Environ. Microbiol. 67:2213-2221 (2001).
DR EMBL; AF339044; AAK53358.1; -.
DR InterPro; IPR001117; Cu-oxidase.
DR Pfam; PF00394; Cu-oxidase; 1.
FT NON_TER 1
FT NON_TER 1
FT SEQUENCE 180 AA; 19681 MW; 438BD2BE09868FE CRC64;
SQ

Query Match
Best Local Similarity 34.1%; Pred. No. 5.2e-08;
Matches 63; Conservative 33; Mismatches 62; Indels 27; Gaps 9;

QY 177 VYHCAVAPVGM---HIANGMYGLILVEPKEGLP-----KVDKEYYVMQGDIFYK-----T 221
DB 1 VYHC--APGGMVPMWHTSGMNGAIWVLPDGLKDEKQGPLTYDKIYVGEOPFYPKDEA 58
QY 223 ---GKYGEQGLQPFDMKEAIR-DAEYVFNKSVGALTGEMNAKAKVGETVRLFVNGG 277
DB 59 GNYKKTETGEAYEDVAKMRILTPHYFNKVGALTDGHALTAAGV--KVLVYHSAQ 116
QY 279 NLTSSEFHVIGEIFDKVHFEKGKGNH-----IQTLLIPAGGAITEFKVDVPGDYVLVD 334
DB 117 NADTRPHLIGGHQDYV---GATGCFKFNPPDLQDETLLIPGTAAGAFYTRQPGVAYAVN 173
QY 334 AIFRA 339
DB 176 NMLIEA 180

RESULT 22
ID Q93TN3 PRELIMINARY; PRT; 180 AA.
AC Q93TN3;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)

```

```

DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Putative dissimilatory nitrite reductase (Fragment).
GN Nitr.
OS Nitrosomonas sp. C-45.
OC Bacteria; Proteobacteria; Betaproteobacteria; Nitrosomonadales;
OC Nitrosomonadaceae; Nitrosomonas.
OX NCBI_Taxid=159353;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-45;
RC MEDLINE=21218534; PubMed=11319103;
RA Casciotti K.L., Ward B.B.;
RT "Dissimilatory nitrite reductase genes from autotrophic ammonia-oxidizing bacteria.";
RL Appl. Environ. Microbiol. 67:2213-2221 (2001).
DR EMBL; AF339047; AAK53361.1; -.
DR InterPro; IPR001117; Cu-oxidase.
DR Pfam; PF00394; Cu-oxidase; 1.
FT NON_TER 1
FT NON_TER 1
FT SEQUENCE 180 AA; 19540 MW; BB7CA048F7CD6D29 CRC64;
SQ

Query Match
Best Local Similarity 35.1%; Pred. No. 7.1e-08;
Matches 65; Conservative 30; Mismatches 63; Indels 27; Gaps 9;

QY 177 VYHCAVAPVGM---HIANGMYGLILVEPKEGLP-----KVDKEYYVMQGDIFYK-----T 221
DB 1 VYHC--APGGMVPMWHTSGMNGAIWVLPDGLKDEKQGPLTYDKIYVGEOPFYPKDEA 58
QY 223 ---GKYGEQGLQPFDMKEAIR-DAEYVFNKSVGALTGEMNAKAKVGETVRLFVNGG 277
DB 59 GNYKKTETGEAYEDVAKMRILTPHYFNKVGALTDGHALTAAGV--KVLVYHSAQ 116
QY 279 NLTSSEFHVIGEIFDKVHFEKGKGNH-----IQTLLIPAGGAITEFKVDVPGDYVLVD 334
DB 117 NADTRPHLIGGHQDYV---GATGCFKFNPPDLQDETLLIPGTAAGAFYTRQPGVAYAVN 173
QY 334 AIFRA 339
DB 176 NMLIEA 180

RESULT 23
ID Q93TN4 PRELIMINARY; PRT; 180 AA.
AC Q93TN4;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Putative dissimilatory nitrite reductase (Fragment).
GN Nitr.
OS Nitrosomonas sp. URW.
OC Bacteria; Proteobacteria; Betaproteobacteria; Nitrosomonadales;
OC Nitrosomonadaceae; Nitrosomonas.
OX NCBI_Taxid=159352;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=URW;
RC MEDLINE=21218534; PubMed=11319103;
RA Casciotti K.L., Ward B.B.;
RT "Dissimilatory nitrite reductase genes from autotrophic ammonia-oxidizing bacteria.";
RL Appl. Environ. Microbiol. 67:2213-2221 (2001).
DR EMBL; AF339046; AAK53360.1; -.
DR InterPro; IPR001117; Cu-oxidase.
DR Pfam; PF00394; Cu-oxidase; 1.
FT NON_TER 1
FT NON_TER 1
FT SEQUENCE 180 AA; 19540 MW; BB7CA048F7CD6D29 CRC64;
SQ

Query Match
Best Local Similarity 35.1%; Pred. No. 7.1e-08;

```





Matches 65; Conservative 24; Mismatches 60; Indels 27; Gaps 8;

QY 185 VGHINANGMGLIVEKEGLP-----KVDKEYVMQGDFT----KCKY-----GE 227  
 DB 3 VPMHVVSGMGAAMVPRREGKKGAGKALYDRAVYIGEHLVPRDEKCKYRYSFGE 62  
 QY 228 OGQAPPD-MEKAREDEYVFNVSVALTGENALKAKVETVRLFEVNGSPNLTSSFHV 286  
 DB 63 AYDDTMEMLMKLI---PTHVFNCKGALTKNAMAIVKQVETV--LVHSQARNDSRPHL 117  
 QY 287 IGEIFDKVHEGEGKGENH---NIQTLLIPAGAAITFEKVDVPGDYLVDHAI 339  
 DB 118 IGHGHVH-VETGKFARPEVDLETWFIKGGSGAALYTRPGIYAVVHNLIEA 172

RESULT 27

Q8RLT1 PRELIMINARY; PRT; 179 AA.

ID Q8RLT1  
 AC Q8RLT1;  
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 21, Last sequence update)  
 DE Putative dissimilatory nitrite reductase (Fragment).  
 GN NTRK.  
 OS Ochrobactrum sp. 3CB4.  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 OC Brucellaceae; Ochrobactrum.  
 NC NCB1\_TaxID=126140;  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=3CB4;  
 RC Song B., Ward B.B.;  
 RA "Nitrite reductase genes in halobenzoate degrading denitrifying  
 RT bacteria";  
 RL Submitted (FE8-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AY078250; AAL82509.1;  
 DR InterPro; IPR001117; Cu-oxidase.  
 DR Pfam; PF00394; Cu-oxidase; 1.  
 FT NON\_TER 1 1  
 FT NON\_TER 179 179  
 SQ SEQUENCE 179 AA; 19587 MW; 021DB57B2A88A4F CRC64;

Query Match 8.5%; Score 223; DB 2; Length 179;  
 Best Local Similarity 36.6%; Pred. No. 1.7e-07;  
 Matches 67; Conservative 23; Mismatches 69; Indels 24; Gaps 8;

QY 177 VYHCAPVGM--HINANGMGLIVEKEGLP-----KVDKEYVMQGDFT----K 222  
 DB 1 VYHC--APGKVPWHVVSGMGAAMVPRREGKKGKALYDRIYVVGEDDFVPRDEN 58  
 QY 223 GKVE---OGQAPPDMEKARE--DAEVVFNVSVALTGENALKAKVETVRLFEVNGSP 278  
 DB 59 GKAKKVPAGDVAEDYVKNMTLTPTVFNVSVALTGENALKAKVETVRLFEVNGSP 116  
 QY 279 NLTSFVIGEIFDKVHEGEGKGENHNI--QTLLIPAGAAITFEKVDVPGDYLVDHAI 336  
 DB 117 NRDRPHLIGHGHDYVATGKFNTPPVDDETWFIKGGSGAALYTRPGIYAVVHNL 176

QY 337 FRA 339  
 DB 177 IEA 179

RESULT 28

Q9K098 PRELIMINARY; PRT; 138 AA.

ID Q9K098  
 AC Q9K098;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Cytochrome, putative.  
 GN NMB0717.  
 OS Neisseria meningitidis (serogroup B).

OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;  
 OC Neisseriaceae; Neisseria.  
 NC NCB1\_TaxID=491;  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=MC58 / Serogroup B;  
 RX MEDLINE=20175755; PubMed=10710307;  
 RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,  
 RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,  
 RA Nelson W.C., Gwin M.L., Deboy R., Peterson J.D., Hickey E.K.,  
 RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,  
 RA Mason T., Ciecko A., Parksey D.S., Blair E., Clifton H., Clark E.B.,  
 RA Cotton M.D., Ulfteback T.R., Khouri H., Qin H., Vamathevan J.,  
 RA Gill J., Scariato V., Masignani V., Pizza M., Grandi G., Sun L.,  
 RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;  
 RT "Complete genome sequence of Neisseria meningitidis serogroup B strain  
 RL MC58";  
 RL Science 287:1809-1815 (2000).  
 DR EMBL; AE002426; AAF41130.1; -;  
 DR HSSP; P04164; 1C52.  
 DR TIGR; NMB0717; -;  
 DR InterPro; IPR000345; Cytochrome\_heme\_bind.  
 DR InterPro; IPR003088; Cytochrome\_heme\_bind.  
 DR Pfam; PF00034; Cytochrome\_c; 1.  
 DR PROSITE; PS00190; CYTOCHROME\_C; 1.  
 KM Complete proteome.  
 SQ SEQUENCE 138 AA; 14715 MW; 947903CE2CBE6F23 CRC64;

Query Match 8.5%; Score 222; DB 16; Length 138;  
 Best Local Similarity 42.7%; Pred. No. 1.3e-07;  
 Matches 44; Conservative 14; Mismatches 45; Indels 0; Gaps 0;

QY 398 GKATYDNCACQHPDGKGVNAPPLANSDYLNADHARAASIVANGSGKITVNGOYE 457  
 DB 35 GKATYDNCACQHPDGKGVNAPPLANSDYLNADHARAASIVANGSGKITVNGOYE 94  
 QY 458 SWPAPATLSDQIANTYITLNSFGNKGQSLADYAKAKKTK 500  
 DB 95 GFWPATATLSDQIANTYITLNSFGNKGQSLADYAKAKKTK 137

RESULT 29

Q9JVA7 PRELIMINARY; PRT; 163 AA.

ID Q9JVA7  
 AC Q9JVA7;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Putative cytochrome C.  
 GN NMA0925.  
 OS Neisseria meningitidis (serogroup A).  
 OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;  
 OC Neisseriaceae; Neisseria.  
 NC NCB1\_TaxID=65699;  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=Z2491 / Serogroup A; Serotype 4A;  
 RX MEDLINE=20222556; PubMed=10761919;  
 RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,  
 RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,  
 RA Davies R.M., Davis P., Devlin K., Feltham T., Hamlin N., Holtroyd S.,  
 RA Jagers K., Leach S., Moule S., Mungall K., Quail M.A.,  
 RA Rajadaram M.A., Rutherford K.M., Simmonds M., Skelton J.,  
 RA Whitehead S., Spratt B.G., Barrall B.G.;  
 RT "Complete DNA sequence of a serogroup A strain of Neisseria  
 RL meningitidis Z2491";  
 RL Nature 404:502-506 (2000).  
 DR EMBL; AL162754; CAB84197.1; -;  
 DR HSSP; P04164; 1C52.  
 DR InterPro; IPR000345; Cytochrome\_heme\_bind.  
 DR InterPro; IPR003088; Cytochrome\_heme\_bind.  
 DR Pfam; PF00034; Cytochrome\_c; 1.  
 DR PROSITE; PS00190; CYTOCHROME\_C; 1.



Db 1 VVHC--APPGMVMHVSGNMGAIWLPREGLDGKALTYDKIYVGEODFYVPRD 58  
QY 223 ---GKYGEGLOFDEMEKARE--DAEYVFNKSGALTGEMAKAVGETVRLFVNGGP 278  
Db 59 GKFKYDAPDADVEDYKVRILTPTTHVFNKSGALTGEMAKAVGETVRLFVNGGP 111  
QY 279 NLTSSFVHIGELFDKVFEGGKGE-----NHNIQTLLIPAGGAITEFVNDVP 326  
Db 112 -----VHSQANRDRPHLKGHDYDWSTGKFTPTPDVDETFTFGAAGAFYTFQGP 166  
QY 327 GDYVLDHAIIRA 339  
Db 167 GIYAVVNHNLIEA 179

RESULT 33  
Q8RT15 PRELIMINARY; PRT; 180 AA.  
AC Q8RT15; 01-JUN-2002 (Tremblrel. 21, Created)  
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)  
DE 01-OCT-2002 (Tremblrel. 22, Last annotation update)  
DR Copper-containing nitrite reductase (Fragment).  
GN N1RK.  
OS Azospirillum dobereineriae.  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodospirillales;  
OC Rhodospirillaceae; Azospirillum.  
OX NCBI\_TaxID=92933;  
RN 11)  
RP SEQUENCE FROM N.A.  
RC STRAIN=GSF 71T;  
RA Roesch C., Mergel A., Bothe H.;  
RT "Biodiversity of denitrifying and dinitrogen fixing bacteria in an  
acid forest soil."  
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AY072263; AL73092.1;  
DR InterPro; IPR001117; Cu-oxidase.  
DR Pfam; PF00394; Cu-oxidase; 1.  
FT NON\_TER 1  
FT NON\_TER 180  
SQ SEQUENCE 180 AA; 19881 MW; 9A451553E7D6EFC CRC64;

Query Match 8.3%; Score 219; DB 2; Length 180;  
Best Local Similarity 33.7%; Pred. No. 3, 1e-07;  
Matches 66; Conservative 26; Mismatches 56; Indels 48; Gaps 10;

QY 176 VYHCAPVGM--HIANGYGLIYVPEKGLP-----KDKKEYVMQGFY---T 221  
Db 1 FVYHC--APPGMIPMHVSGNMGAIWLPREGLDGKALTYDKIYVGEODFYVPRD 58  
QY 222 KGKYG-----OGIOPFDEMEKAREDAEYVFNKSGALTGEMAKAVGETVRL 266  
Db 59 NGKXKYESAGDSYADTVEVNGKGLP-----THVFNKSGALTGEMAKAVGETV 107  
QY 267 ETVRLFVNGGPNLTSSFVHIGELFDKVFEGGKGE---HNIIQTLLIPAGGAITEFV 323  
Db 108 ETV-LFI-HSANDRTPHLKGHDYDWSTGKFTPTPDVDETFTFGAAGAFYTF 164  
QY 324 DVPEDYVLDHAIIRA 339  
Db 165 HOPGIYAVVNHNLIEA 180

RESULT 34  
Q8KPL0 PRELIMINARY; PRT; 172 AA.  
AC Q8KPL0; 01-OCT-2002 (Tremblrel. 22, Created)  
DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)  
DE 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
DR Nitrite reductase (Fragment).  
GN N1RK.  
OS uncultured bacterium.

OC Bacteria; environmental samples.  
OX NCBI\_TaxID=77133;  
RN 11)  
RP SEQUENCE FROM N.A.  
RC MEDLINE=21914100; PubMed=11916709;  
RA Prieme A., Braker G., Tiedje J.M.;  
RT "Diversity of nitrite reductase (nirK and nirS) gene fragments in  
forested upland and wetland soils."  
RL Appl. Environ. Microbiol. 68:1893-1900 (2002).  
DR EMBL; AY121544; AAM94075.1;  
DR InterPro; IPR001117; Cu-oxidase.  
DR Pfam; PF00394; Cu-oxidase; 1.  
FT NON\_TER 1  
FT NON\_TER 172  
SQ SEQUENCE 172 AA; 19273 MW; E95D4BB6B340A544 CRC64;

Query Match 8.2%; Score 216.5; DB 2; Length 172;  
Best Local Similarity 35.2%; Pred. No. 4, 3e-07;  
Matches 62; Conservative 27; Mismatches 60; Indels 27; Gaps 8;

QY 185 VGMHANGVYGLIYVPEKGLP-----KVDKEYVMQGFY---KGYT-----GE 227  
Db 3 VMHVVSGMHGVWVLPREGLDGKALTYDKIYVGEODFYVPRD 58  
QY 228 OGLOPFD--MEKARIDAEYVFNKSGALTGEMAKAVGETVRLFVNGGPNLTSSFV 286  
Db 63 SFADTTEVVRKLI-----PTHVFNKSGALTGEMAKAVGETV--MIVHSQANRDRPHL 117  
QY 287 IGEIFPKVHFEKGKGE---HNIIQTLLIPAGGAITEFVNDVPDGYVLDHAIIRA 339  
Db 118 IGHGHDYV-WETGKFNPNPQKDETFIRGSGAALYTFREPDIYAVVNHNLIEA 172

RESULT 35  
Q8RLT0 PRELIMINARY; PRT; 179 AA.  
AC Q8RLT0; 01-JUN-2002 (Tremblrel. 21, Created)  
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)  
DE 01-OCT-2002 (Tremblrel. 22, Last annotation update)  
DR Putative dissimilatory nitrite reductase (Fragment).  
GN N1RK.  
OS Ochrobactrum sp. 3CB5.  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
OC Brucellaceae; Ochrobactrum.  
OX NCBI\_TaxID=126141;  
RN 11)  
RP SEQUENCE FROM N.A.  
RC STRAIN=3CB5;  
RA Song B., Ward B.B.;  
RT "Nitrite reductase genes in halobenzate degrading denitrifying  
bacteria."  
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AY078251; AAL82510.1;  
DR InterPro; IPR001117; Cu-oxidase.  
DR Pfam; PF00394; Cu-oxidase; 1.  
FT NON\_TER 1  
FT NON\_TER 179  
SQ SEQUENCE 179 AA; 19512 MW; C51DC3A43FEE3A48 CRC64;

Query Match 8.2%; Score 216; DB 2; Length 179;  
Best Local Similarity 36.1%; Pred. No. 4, 9e-07;  
Matches 66; Conservative 23; Mismatches 70; Indels 24; Gaps 8;

QY 177 VYHCAPVGM--HIANGYGLIYVPEKGLP-----KVDKEYVMQGFY---K 222  
Db 1 VVHC--APPGMVMHVSGNMGAIWLPREGLDGKALTYDKIYVGEODFYVPRD 58  
QY 223 GKYGEGLOFDEMEKARE--DAEYVFNKSGALTGEMAKAVGETVRLFVNGGP 278  
Db 59 GKFKYDAPDADVEDYKVRILTPTTHVFNKSGALTGEMAKAVGETVRLFVNGGP 111  
QY 279 NLTSSFVHIGELFDKVFEGGKGE---NHNIQTLLIPAGGAITEFVNDVP 326

Db 117 NRQTRPHLIGHGHDYVWATGKENTPPDVDETFIRGGAAGAAFYTPQGGIYAVVNHNL 176  
 QY 337 FRA 339  
 Db 177 IEA 179

## RESULT 36

Q8RLT2 PRELIMINARY; PRT; 179 AA.

AC Q8RLT2;  
 DT 01-JUN-2002 (T-EMBLrel. 21, Created)  
 DT 01-OCT-2002 (T-EMBLrel. 21, Last sequence update)  
 DT 01-OCT-2002 (T-EMBLrel. 22, Last annotation update)  
 DE Putative dissimilatory nitrite reductase (Fragment).  
 GN NIRK.  
 OS Ochrobactrum sp. 2FBI0.  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 OC Brucellaceae; Ochrobactrum.  
 OX NCB1\_TaxID=126135;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=2FBI0.  
 RA Song B., Ward B.B.;  
 RT "Nitrite reductase genes in halobenzoeate degrading denitrifying  
 bacteria."  
 RT bacteriophage.  
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AY078249; AAL82508.1;  
 DR InterPro; IPR001117; Cu-oxidase.  
 DR Pfam; PF00394; Cu-oxidase; 1.  
 FT NON\_TER 1 1  
 FT SEQUENCE 179 AA; 19606 MW; 021DDB57B2BDA4F CRC64;

Query Match 8.2%; Score 215; DB 2; Length 179;  
 Best Local Similarity 36.1%; Pred. No. 5.8e-07;  
 Matches 66; Conservative 23; Mismatches 70; Indels 24; Gaps 8;

QY 177 VYHCAVAPYGM--HIANGVYGLIVEPKKGLP-----KVDKEYVYMGDPYT-----K 222  
 Db 1 VYHC-APRGWVWVHVSGMAGAIMVLPREGLDGKSKALTYKITYVGEQDPTVPRDEN 58  
 QY 223 GRYGE--OGLQPFDEKAIAR-DAEVYVNSVGLTGENALKAKVGETVRLFVNGNGP 278  
 Db 59 GKXKYEAPGDAVEDTVKVRILTTPHYVNGVGLTGPKATAVGE--KYLIVHSQA 116  
 QY 279 NLTSSHVIGEIFDKVHFEGGKGENNT--QTLIPAGGAITEFKVDVPGDYVLVDHAI 336  
 Db 117 NRQTRPHLIGHGHDYVWATGKENTPPDVDETFIRGGAAGAAFYTPQGGIYAVVNHNL 176  
 QY 337 FRA 339  
 Db 177 IEA 179

## RESULT 37

Q8KPM6 PRELIMINARY; PRT; 172 AA.

AC Q8KPM6;  
 DT 01-OCT-2002 (T-EMBLrel. 22, Created)  
 DT 01-OCT-2002 (T-EMBLrel. 22, Last sequence update)  
 DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)  
 DE Nitrite reductase (Fragment).  
 GN NIRK.  
 OS uncultured bacterium.  
 OC Bacteria; environmental samples.  
 OC NCB1\_TaxID=77133;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=21914100; PubMed=11916709;  
 RA Prieme A., Braker G., Tiedje J.M.;  
 RT "Diversity of nitrite reductase (nirK and nirS) gene fragments in

RT forested upland and wetland soils."  
 RL Appl. Environ. Microbiol. 68:1893-1900 (2002).  
 DR EMBL; AY121527; AAM94058.1;  
 DR InterPro; IPR001117; Cu-oxidase.  
 DR Pfam; PF00394; Cu-oxidase; 1.  
 FT NON\_TER 1 1  
 FT SEQUENCE 172 AA; 19069 MW; B477803B4AFCE50F CRC64;

Query Match 8.1%; Score 213.5; DB 2; Length 172;  
 Best Local Similarity 35.1%; Pred. No. 6.9e-07;  
 Matches 61; Conservative 27; Mismatches 63; Indels 23; Gaps 7;

QY 185 VGMHANGVYGLIVEPKKGLP-----KVDKEYVYMGDPYT-----TKGKYEQGL 230  
 Db 3 VGMHANGVYGLIVEPKKGLP-----KVDKEYVYMGDPYT-----TKGKYEQGL 230  
 QY 231 QEPDEKAIRED--AEVYVNSVGLTGENALKAKVGETVRLFVNGNGPNTSSHFVIG 288  
 Db 62 ESYDETMAVVRGLIPHYVFNKGVSLTGENALKSKVGETV--LIVHSQANRDTXHLIG 119  
 QY 289 EIFDKVHFEGGKGEN--HNIOITLLIPAGGAITEFKVDVPGDYVLVDHAI 339  
 Db 120 GHGDYV-WEQGFAPPAKDLTETWTFIRGGSAGAGLYTFIQGGIYAVVNHNL 172

## RESULT 38

Q8KPM5 PRELIMINARY; PRT; 172 AA.

AC Q8KPM5;  
 DT 01-OCT-2002 (T-EMBLrel. 22, Created)  
 DT 01-OCT-2002 (T-EMBLrel. 22, Last sequence update)  
 DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)  
 DE Nitrite reductase (Fragment).  
 GN NIRK.  
 OS uncultured bacterium.  
 OC Bacteria; environmental samples.  
 OC NCB1\_TaxID=77133;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=21914100; PubMed=11916709;  
 RA Prieme A., Braker G., Tiedje J.M.;  
 RT "Diversity of nitrite reductase (nirK and nirS) gene fragments in  
 forested upland and wetland soils."  
 RL Appl. Environ. Microbiol. 68:1893-1900 (2002).  
 DR EMBL; AY121528; AAM94059.1;  
 DR InterPro; IPR001117; Cu-oxidase.  
 DR Pfam; PF00394; Cu-oxidase; 1.  
 FT NON\_TER 1 1  
 FT SEQUENCE 172 AA; 19275 MW; 9DCC6990239F1C0A CRC64;

Query Match 8.1%; Score 211.5; DB 2; Length 172;  
 Best Local Similarity 34.7%; Pred. No. 9.4e-07;  
 Matches 61; Conservative 27; Mismatches 61; Indels 27; Gaps 8;

QY 185 VGMHANGVYGLIVEPKKGLP-----KVDKEYVYMGDPYT-----KGKYEQGL 227  
 Db 3 VGMHANGVYGLIVEPKKGLP-----KVDKEYVYMGDPYT-----KGKYEQGL 227  
 QY 228 QGLQPFDEKAIRED--AEVYVNSVGLTGENALKAKVGETVRLFVNGNGPNTSSHFV 286  
 Db 63 SYADTTEVVRKI---PHTVYVNGVGLTGENALKSKVGETVRL--HSQANRDTXHLIG 117  
 QY 287 ICEIFDKVHFEGGKGEN--HNIOITLLIPAGGAITEFKVDVPGDYVLVDHAI 339  
 Db 118 IGHGHDYV-WEAGKKNPPOKDLTETWTFVFGSAGAGALYTFRBPDIYAVVNHNL 172

## RESULT 39

Q8KPK9 PRELIMINARY; PRT; 172 AA.

AC Q8KPK9;  
 QY 172 IEA 179

DT 01-OCT-2002 (TEMBLrel. 22, Created)  
 DT 01-OCT-2002 (TEMBLrel. 22, Last sequence update)  
 DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)  
 DE Nitrite reductase (Fragment).  
 GN NITRK.  
 OS uncultured bacterium.  
 OC Bacteria; environmental samples.  
 OX NCBI\_TaxID=77133;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21914100; PubMed=11916709;  
 RA Prieme A., Braker G., Tiedje J.M.;  
 RT "Diversity of nitrite reductase (nitr and nirs) gene fragments in  
 forested upland and wetland soils."  
 RL Appl. Environ. Microbiol. 68:1893-1900(2002).  
 DR EMBL: AY121545; AAM94076.1; -  
 DR InterPro: IPR001117; Cu-oxidase.  
 DR Pfam: PF00394; Cu-oxidase; 1.  
 FT NON\_TER 1  
 FT 172 1  
 SQ SEQUENCE 172 AA; 19137 MW; 06E260DBA5B345 CRC64;

Query Match 8.1%; Score 211.5; DB 2; Length 172;  
 Best Local Similarity 34.7%; Pred. No. 9.4e-07;  
 Matches 61; Conservative 27; Mismatches 61; Indels 27; Gaps 8;

QY 185 VGMHANGMYGLIVERKGLP-----KVDKEYVMGGDFYT---KGKY-----GE 227  
 DB 3 VPMHVVSGMHGTWVLPKDLKDAEGKALRYDKIFYIGENDLYVPKDENGKYTYETIGE 62  
 QY 228 OGLQFPD-MEKAIRDAEYVFNQSVGALTGENALKAKVETVRLFVNGGPNLTSSFHV 286  
 DB 63 SYADTTEVVRKLI---PTHVFNQSVGLTGKAMTSKVGETVMLV--HSQANRDTTRPHL 117  
 QY 287 TGEIFDKVHEGGKGEN--HNIOITLIPAGAAITTEFKVDVPGDYLVDHAIFFA 339  
 DB 118 IGHGDHV-WEAGKFNPNPKDLETFWIRGSGAGALYTFREPGIYAYVNHNLIEA 172

## RESULT 40

Q8KPM9 PRELIMINARY; PRT; 172 AA.  
 AC Q8KPM9;  
 DT 01-OCT-2002 (TEMBLrel. 22, Created)  
 DT 01-OCT-2002 (TEMBLrel. 22, Last sequence update)  
 DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)  
 DE Nitrite reductase (Fragment).  
 GN NITRK.  
 OS uncultured bacterium.  
 OC Bacteria; environmental samples.  
 OX NCBI\_TaxID=77133;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21914100; PubMed=11916709;  
 RA Prieme A., Braker G., Tiedje J.M.;  
 RT "Diversity of nitrite reductase (nitr and nirs) gene fragments in  
 forested upland and wetland soils."  
 RL Appl. Environ. Microbiol. 68:1893-1900(2002).  
 DR EMBL: AY121521; AAM94052.1; -  
 DR InterPro: IPR001117; Cu-oxidase.  
 DR Pfam: PF00394; Cu-oxidase; 1.  
 FT NON\_TER 1  
 FT 172 1  
 SQ SEQUENCE 172 AA; 19146 MW; 1B490D6A18695C45 CRC64;

Query Match 8.1%; Score 211.5; DB 2; Length 172;  
 Best Local Similarity 34.1%; Pred. No. 9.4e-07;  
 Matches 60; Conservative 28; Mismatches 61; Indels 27; Gaps 8;

QY 185 VGMHANGMYGLIVERKGLP-----KVDKEYVMGGDFYT---KGKY-----GE 227  
 DB 3 VPMHVVSGMHGTWVLPKDLKDAEGKALRYDKIFYIGENDLYVPKDENGKYTYETIGE 62

QY 228 OGLQFPD-MEKAIRDAEYVFNQSVGALTGENALKAKVETVRLFVNGGPNLTSSFHV 286  
 DB 63 SYADTTEVVRKLI---PTHVFNQSVGLTGKAMTSKVGETVMLV--HSQANRDTTRPHL 117  
 QY 287 TGEIFDKVHEGGKGEN--HNIOITLIPAGAAITTEFKVDVPGDYLVDHAIFFA 339  
 DB 118 IGHGDHV-WEAGKFNPNPKDLETFWIRGSGAGAAIYTFREPGIYAYVNHNLIEA 172

Search completed: August 27, 2003, 18:31:55  
 Job time : 106 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 27, 2003, 18:33:47 ; Search time 104 Seconds  
(without alignments)  
1245.600 Million cell updates/sec

Title: US-10-088-045-2

Perfect score: 502  
Sequence: 1 MSKPLIKITTLICALSALML.....NKGGLSADDVAKAKKTKPN 502

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 830525 seqs, 258052604 residues

Word size : 0

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database :

SPTREMBL\_23:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriophage:\*  
17: sp\_archaeop:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	34	6.8	386	16 Q9JTB8	Q9JTB8 neisseria m
2	34	6.8	390	16 Q9JTE1	Q9JTE1 neisseria m
3	19	3.8	510	16 Q8XPY3	Q8XPY3 rai
4	16	3.2	159	2 Q8GBL8	Q8GBL8 uncultured
5	12	2.4	94	14 Q8BFN8	Q8BFN8 uncultured
6	12	2.4	95	14 Q8BFP2	Q8BFP2 uncultured
7	12	2.4	99	14 Q8BFL7	Q8BFL7 uncultured
8	12	2.4	102	14 Q8BFN9	Q8BFN9 uncultured
9	12	2.4	102	14 Q8BFN5	Q8BFN5 uncultured
10	12	2.4	102	14 Q8BFN3	Q8BFN3 uncultured
11	12	2.4	102	14 Q8BFN0	Q8BFN0 uncultured
12	12	2.4	102	14 Q8BFM8	Q8BFM8 uncultured
13	12	2.4	102	14 Q8BFM2	Q8BFM2 uncultured
14	12	2.4	102	14 Q8BFL4	Q8BFL4 uncultured
15	12	2.4	102	14 Q8BFX9	Q8BFX9 uncultured
16	12	2.4	102	14 Q8BFP6	Q8BFP6 uncultured

17	12	2.4	103	14 Q8BFP0	Q8BFP0 uncultured
18	12	2.4	172	2 Q87751	Q87751 hypomicrob
19	12	2.4	142	2 Q8KPL1	Q8KPL1 uncultured
20	12	2.4	360	2 Q9AQO7	Q9AQO7 alcaligenes
21	12	2.4	486	2 Q8KKH4	Q8KKH4 hypomicrob
22	11	2.2	102	14 Q8BFL0	Q8BFL0 uncultured
23	11	2.2	102	14 Q8BFL6	Q8BFL6 uncultured
24	10	2.0	361	1 Q9P9H9	Q9P9H9 haloarcula
25	9	1.8	172	2 Q8KPL6	Q8KPL6 uncultured
26	9	1.8	428	16 Q8CXH7	Q8CXH7 oceanobacil
27	8	1.6	134	16 Q8X8W0	Q8X8W0 escherichia
28	8	1.6	171	2 Q8KRL1	Q8KRL1 uncultured
29	8	1.6	172	2 Q8KPL6	Q8KPL6 uncultured
30	8	1.6	252	17 Q8TSG4	Q8TSG4 methanosarc
31	8	1.6	252	17 Q8PVL5	Q8PVL5 methanosarc
32	8	1.6	255	10 Q9FFM7	Q9FFM7 arabidopsis
33	8	1.6	321	16 Q8EUC9	Q8EUC9 mycoplasma
34	8	1.6	360	2 Q86601	Q86601 alcaligenes
35	8	1.6	483	16 Q9KLC4	Q9KLC4 neisseria m
36	8	1.6	494	16 Q9JX77	Q9JX77 neisseria m
37	8	1.6	509	16 Q8H112	Q8H112 rhizobium 1
38	8	1.6	514	10 Q9AV50	Q9AV50 oryza sativ
39	8	1.6	634	10 Q9LZB8	Q9LZB8 arabidopsis
40	8	1.6	915	5 Q93538	Q93538 caenorhabdi
41	8	1.6	995	16 Q866F8	Q866F8 bifidobacte
42	7	1.4	61	16 Q8KE17	Q8KE17 chlorobium
43	7	1.4	77	9 Q9MCT9	Q9MCT9 bacterioph
44	7	1.4	90	5 Q97102	Q97102 drosophila
45	7	1.4	98	16 Q8U8P9	Q8U8P9 agrobacteri
46	7	1.4	101	10 Q8GVQ4	Q8GVQ4 oryza sativ
47	7	1.4	102	14 Q8BFL1	Q8BFL1 uncultured
48	7	1.4	102	14 Q8BFL6	Q8BFL6 uncultured
49	7	1.4	102	14 Q8BFL9	Q8BFL9 uncultured
50	7	1.4	102	14 Q8BFL5	Q8BFL5 uncultured
51	7	1.4	102	14 Q8BFL4	Q8BFL4 uncultured
52	7	1.4	102	14 Q8BFL5	Q8BFL5 uncultured
53	7	1.4	102	14 Q8BFL2	Q8BFL2 uncultured
54	7	1.4	102	14 Q8BFL2	Q8BFL2 uncultured
55	7	1.4	103	11 Q8VDZ2	Q8VDZ2 mus muscul
56	7	1.4	104	2 Q53039	Q53039 rhodococcus
57	7	1.4	104	15 Q76163	Q76163 human immu
58	7	1.4	112	2 Q9L477	Q9L477 uncultured
59	7	1.4	112	2 Q9L484	Q9L484 uncultured
60	7	1.4	112	2 Q9L483	Q9L483 uncultured
61	7	1.4	112	2 Q9L478	Q9L478 uncultured
62	7	1.4	112	2 Q9L480	Q9L480 uncultured
63	7	1.4	118	11 Q923G4	Q923G4 rattus norv
64	7	1.4	133	10 Q41939	Q41939 arabidopsis
65	7	1.4	135	10 Q43037	Q43037 populus del
66	7	1.4	139	16 Q8FOT3	Q8FOT3 leptospira
67	7	1.4	142	2 Q87684	Q87684 alcaligenes
68	7	1.4	142	2 Q87686	Q87686 achromobact
69	7	1.4	142	2 Q9R801	Q9R801 achromobact
70	7	1.4	157	2 Q9X4L1	Q9X4L1 paracoccus
71	7	1.4	159	2 Q8GBL3	Q8GBL3 uncultured
72	7	1.4	160	10 Q39454	Q39454 coriulus ave
73	7	1.4	161	16 Q92HF3	Q92HF3 rickettsia
74	7	1.4	165	16 Q8CU81	Q8CU81 staphylococ
75	7	1.4	165	16 Q8CU70	Q8CU70 staphylococ
76	7	1.4	167	5 Q9VZ18	Q9VZ18 streptophila
77	7	1.4	168	2 Q87591	Q87591 streptophila
78	7	1.4	169	2 Q8KPL0	Q8KPL0 uncultured
79	7	1.4	169	11 Q8C329	Q8C329 mus muscul
80	7	1.4	170	5 Q817G2	Q817G2 caenorhabdi
81	7	1.4	170	16 Q8XUF5	Q8XUF5 rai
82	7	1.4	172	2 Q9R363	Q9R363 pseudomonas
83	7	1.4	172	2 Q9WV18	Q9WV18 pseudomonas
84	7	1.4	172	2 Q8KPK0	Q8KPK0 uncultured
85	7	1.4	172	2 Q9R433	Q9R433 pseudomonas
86	7	1.4	172	2 Q8KPL1	Q8KPL1 uncultured
87	7	1.4	172	2 Q8KPL0	Q8KPL0 uncultured
88	7	1.4	172	2 Q9RBY3	Q9RBY3 pseudomonas
89	7	1.4	172	2 Q8KPM7	Q8KPM7 uncultured

90	7	1.4	172	2	Q9R434	Q9r434 pseudomonas	163	7	1.4	268	2	Q69346	Q69346 pseudomonas
91	7	1.4	172	2	Q9R450	Q9r450 pseudomonas	164	7	1.4	268	2	Q9ERT3	Q9ert3 pseudomonas
92	7	1.4	172	2	Q8RPJ7	Q8rpj7 uncultured	165	7	1.4	268	2	Q9FAF9	Q9faf9 pseudomonas
93	7	1.4	172	2	Q9R449	Q9r449 pseudomonas	166	7	1.4	269	2	Q9FAL1	Q9fal1 pseudomonas
94	7	1.4	172	2	Q9RBV2	Q9rbv2 pseudomonas	167	7	1.4	269	2	Q9FAL6	Q9fal6 pseudomonas
95	7	1.4	172	2	Q9R9F5	Q9r9f5 pseudomonas	168	7	1.4	269	2	Q9FAL2	Q9fal2 pseudomonas
96	7	1.4	172	2	Q9R2S2	Q9r2s2 pseudomonas	169	7	1.4	269	2	Q9ERT6	Q9ert6 pseudomonas
97	7	1.4	172	2	Q9R9F8	Q9r9f8 pseudomonas	170	7	1.4	269	2	Q9FAL1	Q9fal1 pseudomonas
98	7	1.4	173	2	Q8KPK3	Q8kpk3 uncultured	171	7	1.4	269	2	Q9ERT7	Q9ert7 pseudomonas
99	7	1.4	173	2	Q8RTI4	Q8rti4 azospirillum	172	7	1.4	269	2	Q9ERT1	Q9ert1 pseudomonas
100	7	1.4	173	16	Q8XY56	Q8xy56 ralestonia s	173	7	1.4	269	2	Q9FAH8	Q9fah8 pseudomonas
101	7	1.4	174	2	Q9WJ1	Q9wj1 pseudomonas	174	7	1.4	269	2	Q9FAH5	Q9fah5 pseudomonas
102	7	1.4	175	16	Q9ZIT3	Q9zit3 rickettsia	175	7	1.4	269	2	Q9FAK7	Q9fak7 pseudomonas
103	7	1.4	179	2	Q8RLS6	Q8rls6 pseudomonas	176	7	1.4	269	2	Q9FAH9	Q9fah9 pseudomonas
104	7	1.4	179	2	Q93TN1	Q93tn1 nitrosomona	177	7	1.4	269	2	Q9ERT8	Q9ert8 pseudomonas
105	7	1.4	180	2	Q8RTI5	Q8rti5 azospirillum	178	7	1.4	269	2	Q9ERT8	Q9ert8 pseudomonas
106	7	1.4	180	2	Q93TN6	Q93tn6 nitrosomona	179	7	1.4	269	2	Q9ERT8	Q9ert8 pseudomonas
107	7	1.4	180	2	Q93TN3	Q93tn3 nitrosomona	180	7	1.4	269	2	Q9ERT7	Q9ert7 pseudomonas
108	7	1.4	180	2	Q93TN2	Q93tn2 nitrosomona	181	7	1.4	269	2	Q69345	Q69345 pseudomonas
109	7	1.4	180	2	Q93TN4	Q93tn4 nitrosomona	182	7	1.4	269	2	Q9FAJ6	Q9faj6 pseudomonas
110	7	1.4	183	5	Q9W10	Q9w10 dtrosophila	183	7	1.4	269	2	Q9ERT9	Q9ert9 pseudomonas
111	7	1.4	185	5	Q9VIZ1	Q9vizi dtrosophila	184	7	1.4	269	2	Q69344	Q69344 pseudomonas
112	7	1.4	187	16	Q8KEZ8	Q8kez8 chlorobium	185	7	1.4	269	2	Q9ERT2	Q9ert2 pseudomonas
113	7	1.4	188	10	Q9FVY0	Q9fvy0 oryza sativ	186	7	1.4	269	2	Q9FAI2	Q9fai2 pseudomonas
114	7	1.4	188	10	Q8H8Y9	Q8h8y9 oryza sativ	187	7	1.4	269	2	Q9ERT8	Q9ert8 pseudomonas
115	7	1.4	191	16	Q8ZFS5	Q8zfs5 yersinia pe	188	7	1.4	269	2	Q9FAG1	Q9fag1 pseudomonas
116	7	1.4	191	5	Q23065	Q23065 caenorhabdi	189	7	1.4	269	2	Q9ERT5	Q9ert5 pseudomonas
117	7	1.4	193	16	Q8XET8	Q8xet8 salmonella	190	7	1.4	269	2	Q9FAL0	Q9fal0 pseudomonas
118	7	1.4	193	16	Q8XDR5	Q8xd5 escherichia	191	7	1.4	269	2	Q69343	Q69343 pseudomonas
119	7	1.4	195	2	Q54673	Q54673 lactococcus	192	7	1.4	269	2	Q9FAJ2	Q9faj2 pseudomonas
120	7	1.4	196	2	Q8GC69	Q8gc69 escherichia	193	7	1.4	269	2	Q9FAK9	Q9fak9 pseudomonas
121	7	1.4	196	2	Q8GC68	Q8gc68 escherichia	194	7	1.4	269	2	Q9FAJ0	Q9faj0 pseudomonas
122	7	1.4	200	2	Q8VNR7	Q8vnr7 glucinobact	195	7	1.4	269	2	Q9EUB0	Q9eub0 pseudomonas
123	7	1.4	201	2	Q50132	Q50132 mycobacteri	196	7	1.4	269	2	Q9FAI4	Q9fai4 pseudomonas
124	7	1.4	202	2	Q54675	Q54675 lactococcus	197	7	1.4	270	2	Q69339	Q69339 pseudomonas
125	7	1.4	202	16	Q9S2G6	Q9s2g6 streptomyce	198	7	1.4	270	2	Q9FAK6	Q9fak6 pseudomonas
126	7	1.4	208	2	Q9AJ86	Q9aj86 lactococcus	199	7	1.4	271	2	Q9FAK8	Q9fak8 pseudomonas
127	7	1.4	209	2	Q9AC98	Q9ac98 lactococcus	200	7	1.4	278	16	Q8VYX1	Q8vyx1 anabaena sp
128	7	1.4	209	2	Q93K29	Q93k29 lactococcus	201	7	1.4	278	16	Q8NQD7	Q8nd7 corynebacte
129	7	1.4	210	2	Q9JN26	Q9jnz6 lactococcus	202	7	1.4	284	16	Q8G264	Q8g264 brucella su
130	7	1.4	210	2	Q9REY1	Q9rey1 lactococcus	203	7	1.4	285	2	Q52264	Q52264 pseudomonas
131	7	1.4	211	10	Q9SST3	Q9sse3 arabidopsis	204	7	1.4	289	10	Q8SBA6	Q8sba6 oryza sativ
132	7	1.4	211	5	Q9S5J3	Q9s5j3 arabidopsi	205	7	1.4	293	5	P91356	P91356 caenorhabdi
133	7	1.4	212	5	P91355	P91355 caenorhabdi	206	7	1.4	293	16	Q9C1J3	Q9c1j3 lactococcus
134	7	1.4	213	2	Q8LWAI	Q8lwai streptomyce	207	7	1.4	296	16	Q87127	Q87127 pseudomonas
135	7	1.4	214	2	Q8LWAI	Q8lwai lactococcus	208	7	1.4	298	4	Q9NZE1	Q9nze1 homo sapien
136	7	1.4	215	10	Q43719	Q43719 lycopersico	209	7	1.4	299	4	Q8N4S8	Q8nas8 homo sapien
137	7	1.4	216	5	Q20601	Q20601 caenorhabdi	210	7	1.4	299	10	Q9SLV2	Q9slv2 arabidopsis
138	7	1.4	218	10	Q8SID3	Q8sid3 oryza sativ	211	7	1.4	299	16	Q8U8U3	Q8u8u3 agrobacteri
139	7	1.4	218	16	Q8XB27	Q8xb27 escherichia	212	7	1.4	299	16	Q8U9Y6	Q8u9y6 agrobacteri
140	7	1.4	222	1	Q9UMN5	Q9umn5 methanother	213	7	1.4	300	5	Q9V7H9	Q9v7h9 dtrosophila
141	7	1.4	228	5	Q9VB99	Q9vb99 dtrosophila	214	7	1.4	300	16	Q9RK57	Q9rk57 streptomyce
142	7	1.4	228	10	Q9FVL8	Q9fvl8 nicotiana a	215	7	1.4	301	5	Q8IHK7	Q8ihk7 habesia div
143	7	1.4	230	15	Q9IE91	Q9ie91 human immun	216	7	1.4	305	16	Q8UD20	Q8ud20 agrobacteri
144	7	1.4	232	16	Q8KAW5	Q8kaws chlorobium	217	7	1.4	309	16	Q9BHV4	Q9bhv4 campylobact
145	7	1.4	239	12	Q9DUQ4	Q9duq4 rice yellow	218	7	1.4	309	16	Q8CR55	Q8cr55 staphylococ
146	7	1.4	245	16	Q9CDU7	Q9cdt7 lactococcus	219	7	1.4	310	5	Q8IHK4	Q8ihk4 habesia div
147	7	1.4	247	10	Q8LIG3	Q8lig3 oryza sativ	220	7	1.4	312	2	P71214	P71214 escherichia
148	7	1.4	248	5	Q8IM07	Q8imq7 dtrosophila	221	7	1.4	312	2	Q8S6Z5	Q8s6z5 escherichia
149	7	1.4	251	16	Q8XXV6	Q8xxv6 ralestonia s	222	7	1.4	312	16	Q8XC86	Q8xc86 escherichia
150	7	1.4	255	16	Q9PCQ2	Q9pcq2 xylella fas	223	7	1.4	313	2	Q9RZ26	Q9rzt6 bordetella
151	7	1.4	257	2	Q9PF40	Q9pf40 stigmatella	224	7	1.4	313	10	Q9LQ88	Q9lqb8 arabidopsis
152	7	1.4	258	10	Q9AUD0	Q9aud0 oryza sativ	225	7	1.4	315	16	Q9PRJ1	Q9prj1 ureaplasma
153	7	1.4	258	17	Q9HHR7	Q9hhz7 halobacteri	226	7	1.4	318	16	Q8KD57	Q8kds7 chlorobium
154	7	1.4	261	16	Q9HY27	Q9hy27 pseudomonas	227	7	1.4	320	16	Q8IFQ4	Q8ifq4 brucella me
155	7	1.4	264	2	Q8GE45	Q8ge45 heliobacilli	228	7	1.4	321	16	Q8PFQ4	Q8pfq4 xanthomonas
156	7	1.4	267	5	Q9VSU7	Q9vsu7 dtrosophila	229	7	1.4	322	10	Q9SOL1	Q9sol1 musa acumin
157	7	1.4	268	2	Q9FAJ8	Q9faj8 pseudomonas	230	7	1.4	322	16	Q97DS2	Q97ds2 clostridium
158	7	1.4	268	2	Q9FAK2	Q9fak2 pseudomonas	231	7	1.4	323	4	Q96102	Q96102 homo sapien
159	7	1.4	268	2	Q9FAK1	Q9fak1 pseudomonas	232	7	1.4	323	16	Q9A7C4	Q9atc4 caulobacter
160	7	1.4	268	2	Q9ERT0	Q9ert0 pseudomonas	233	7	1.4	326	10	Q9ZPK2	Q9zpk2 pium sativ
161	7	1.4	268	2	Q9FAK3	Q9fak3 pseudomonas	234	7	1.4	326	10	Q9SD22	Q9sd22 glycine max
162	7	1.4	268	2	Q9FAK3	Q9fak3 pseudomonas	235	7	1.4	326	10	Q40316	Q40316 medicago sa



236	7	1.4	327	10	048602	048602 glycine max	309	7	1.4	512	2	0936C7	0936C7 streptomyce
237	7	1.4	328	4	043858	043858 homo sapien	310	7	1.4	512	4	08N5C0	08N5C0 homo sapien
238	7	1.4	331	2	044031	044031 ralstonia m	311	7	1.4	513	11	091V97	091V97 mus musculus
239	7	1.4	331	5	081HK2	081HK2 babesia div	312	7	1.4	513	16	08E805	08E805 shewanella
240	7	1.4	334	17	028427	028427 archaeoglob	313	7	1.4	516	16	0970P4	0970P4 streptococ
241	7	1.4	338	16	097H52	097H52 clostridium	314	7	1.4	519	10	08W0H5	08W0H5 oryza sativ
242	7	1.4	338	16	08D532	08D532 vibrio vuln	315	7	1.4	520	4	09UQ43	09UQ43 homo sapien
243	7	1.4	340	10	022317	022317 musa acum	316	7	1.4	521	11	060437	060437 cricetidae
244	7	1.4	341	4	081W73	081W73 homo sapien	317	7	1.4	521	11	09D4R3	09D4R3 streptococ
245	7	1.4	341	11	099L01	099L01 mus musculu	318	7	1.4	521	16	08DPR2	08DPR2 homo sapien
246	7	1.4	341	16	0917A7	0917A7 pseudomonas	319	7	1.4	522	4	096CR7	096CR7 mus musculu
247	7	1.4	341	16	09AB77	09AB77 caulobacter	320	7	1.4	522	11	0923W0	0923W0 mus musculu
248	7	1.4	343	4	096A55	096A55 homo sapien	321	7	1.4	522	11	0923H3	0923H3 ractus norv
249	7	1.4	344	4	09NS73	09NS73 homo sapien	322	7	1.4	524	11	09D2N5	09D2N5 mus musculu
250	7	1.4	344	4	09BS93	09BS93 mus sapien	323	7	1.4	528	2	093R40	093R40 colwellia m
251	7	1.4	349	11	08K1A7	08K1A7 mus musculu	324	7	1.4	529	16	098CJ7	098CJ7 rhizobium l
252	7	1.4	349	11	09JTC4	09JTC4 mus musculu	325	7	1.4	535	11	08BKX1	08BKX1 mus musculu
253	7	1.4	352	2	052378	052378 pseudomonas	326	7	1.4	538	16	092XB1	092XB1 rhizobium m
254	7	1.4	359	5	020880	020880 caenorhabdi	327	7	1.4	543	5	020720	020720 caenorhabdi
255	7	1.4	363	16	08XQ33	08XQ33 ralstonia s	328	7	1.4	544	5	027659	027659 giardia lam
256	7	1.4	363	16	08KBE1	08KBE1 chlorobium	329	7	1.4	545	16	08RHN0	08RHN0 fusobacteri
257	7	1.4	373	9	09AZR7	09AZR7 bacterioph	330	7	1.4	546	17	09HLE7	09HLE7 thermoplaam
258	7	1.4	373	16	09CFT0	09CFT0 lactococcus	331	7	1.4	547	2	P74913	P74913 treponema d
259	7	1.4	373	17	09V014	09V014 pyrococcus	332	7	1.4	548	4	09BWZ5	09BWZ5 homo sapien
260	7	1.4	376	16	09YB82	09YB82 bruceella me	333	7	1.4	552	4	09UQB8	09UQB8 homo sapien
261	7	1.4	376	16	08FX29	08FX29 bruceella su	334	7	1.4	553	4	081Z85	081Z85 homo sapien
262	7	1.4	379	2	085863	085863 sphingomona	335	7	1.4	554	4	09YJF0	09YJF0 homo sapien
263	7	1.4	380	17	0971D3	0971D3 sulfolobus	336	7	1.4	555	5	0961R9	0961R9 drosophila
264	7	1.4	384	17	09HKY2	09HKY2 thermoplaam	337	7	1.4	558	4	08NA06	08NA06 homo sapien
265	7	1.4	388	16	08EVA2	08EVA2 mycoplasma	338	7	1.4	565	4	081WG6	081WG6 homo sapien
266	7	1.4	389	16	08EVA0	08EVA0 mycoplasma	339	7	1.4	567	16	09ZME2	09ZME2 helicobacte
267	7	1.4	391	3	09C1C2	09C1C2 glomerella	340	7	1.4	567	16	08XS30	08XS30 ralstonia s
268	7	1.4	391	16	09XZ23	09XZ23 thermocoga	341	7	1.4	571	17	08TXB7	08TXB7 methanopyru
269	7	1.4	393	10	09LMD7	09LMD7 camellia si	342	7	1.4	576	4	096W93	096W93 homo sapien
270	7	1.4	393	10	09AT55	09AT55 elaeagnus u	343	7	1.4	583	16	08RFT9	08RFT9 fusobacteri
271	7	1.4	393	10	09AT56	09AT56 elaeagnus u	344	7	1.4	586	4	09NZ29	09NZ29 homo sapien
272	7	1.4	393	10	09AT56	09AT56 elaeagnus u	345	7	1.4	587	4	09H859	09H859 ractus norv
273	7	1.4	396	10	09FPL6	09FPL6 suadeta mari	346	7	1.4	588	11	09D604	09D604 mus musculu
274	7	1.4	398	4	08N6H2	08N6H2 oryza sativ	347	7	1.4	588	11	08BTW8	08BTW8 mus musculu
275	7	1.4	399	4	081VQ3	081VQ3 homo sapien	348	7	1.4	594	11	08BD21	08BD21 mus musculu
276	7	1.4	399	10	08WDM3	08WDM3 oryza sativ	349	7	1.4	594	11	08BD21	08BD21 mus musculu
277	7	1.4	399	16	08ESU4	08ESU4 oceanobacil	350	7	1.4	595	10	0946Y5	0946Y5 chlamydomon
278	7	1.4	403	5	09U250	09U250 caenorhabdi	351	7	1.4	600	5	09V614	09V614 drosophila
279	7	1.4	411	3	09C196	09C196 aspergillus	352	7	1.4	608	16	09HZV4	09HZV4 pseudomonas
280	7	1.4	412	12	09LTM4	09LTM4 tupiaia hep	353	7	1.4	614	17	097B30	097B30 thermoplaam
281	7	1.4	413	5	08SS43	08SS43 entcephalito	354	7	1.4	616	2	087812	087812 pseudomonas
282	7	1.4	419	4	043611	043611 homo sapien	355	7	1.4	617	2	09EVD8	09EVD8 pseudomonas
283	7	1.4	419	10	09LUT4	09LUT4 arabidopsis	356	7	1.4	617	2	09EVB2	09EVB2 pseudomonas
284	7	1.4	420	16	0982K3	0982K3 rhizobium l	357	7	1.4	617	2	09RBY1	09RBY1 pseudomonas
285	7	1.4	420	16	09RUE8	09RUE8 deinococcus	358	7	1.4	617	2	09X3C0	09X3C0 pseudomonas
286	7	1.4	429	10	08GTL5	08GTL5 carica papa	359	7	1.4	663	16	09KR77	09KR77 vibrio chol
287	7	1.4	431	2	08GLD8	08GLD8 bacillus th	360	7	1.4	672	5	09VMU5	09VMU5 drosophila
288	7	1.4	438	2	052582	052582 staphylococ	361	7	1.4	714	16	09A5S3	09A5S3 caulobacter
289	7	1.4	438	16	09KAH8	09KAH8 bacillus ha	362	7	1.4	714	16	09RZD1	09RZD1 amycolalope
290	7	1.4	438	16	099VC0	099VC0 staphylococ	363	7	1.4	722	2	09F164	09F164 notophthalm
291	7	1.4	438	16	08NXX8	08NXX8 staphylococ	364	7	1.4	744	13	091149	091149 pseudomonas
292	7	1.4	447	5	020703	020703 caenorhabdi	365	7	1.4	747	2	09ZAX6	09ZAX6 lactococcus
293	7	1.4	450	2	059658	059658 pelobacter	366	7	1.4	747	2	09CIT3	09CIT3 lactococcus
294	7	1.4	451	2	045695	045695 butkholderi	367	7	1.4	764	2	08KFW9	08KFW9 saccharopol
295	7	1.4	455	16	08U7R7	08U7R7 agrobacteri	368	7	1.4	779	16	08YNN0	08YNN0 anabaena sp
296	7	1.4	466	3	000100	000100 aspergillus	369	7	1.4	789	17	08TXR3	08TXR3 methanopyru
297	7	1.4	472	2	08KZS7	08KZS7 acetobacter	370	7	1.4	800	12	09WAS5	09WAS5 zucchini ye
298	7	1.4	472	2	057536	057536 acetobacter	371	7	1.4	800	12	09WAS4	09WAS4 zucchini ye
299	7	1.4	473	16	08CJS9	08CJS9 streptomyce	372	7	1.4	804	2	08GCV4	08GCV4 lactobacill
300	7	1.4	479	16	08FTW0	08FTW0 corynebacte	373	7	1.4	815	13	P87376	P87376 tritriturus ca
301	7	1.4	491	5	096638	096638 trypanosoma	374	7	1.4	844	16	025856	025856 helicobacte
302	7	1.4	494	16	08D7Z7	08D7Z7 vibrio vuln	375	7	1.4	849	16	09ZJW1	09ZJW1 african swi
303	7	1.4	495	16	092UJ3	092UJ3 rhizobium m	376	7	1.4	868	12	065216	065216 african swi
304	7	1.4	501	16	09A496	09A496 caulobacter	377	7	1.4	869	5	081G55	081G55 caenorhabdi
305	7	1.4	509	12	08QNJ9	08QNJ9 ectocarpus	378	7	1.4	878	16	08BNK4	08BNK4 mycoplasma
306	7	1.4	510	4	09BVG6	09BVG6 homo sapien	379	7	1.4	884	11	08BNK6	08BNK6 mus musculu
307	7	1.4	511	2	09K4U0	09K4U0 mycobacteri	380	7	1.4	884	15	08Q716	08Q716 human immun
308	7	1.4	511	5	09W2H0	09W2H0 drosophila	381	7	1.4	889	2	08RMZ8	08RMZ8 azotobacter

382	7	1.4	923	10	049003	049003	avena sativ	455	6	1.2	57	16	096MG9	096mg rhizobium 1
383	7	1.4	955	12	092625	092625	salmonella	456	6	1.2	57	16	08NPJ7	08npj7 corynebacte
384	7	1.4	970	16	091TM6	091tm6	tupia herp	457	6	1.2	58	12	09DI08	09di08 hepaticis c
385	7	1.4	988	11	08C035	08c035	mus musculu	458	6	1.2	58	13	013084	013084 discipye o
386	7	1.4	1012	11	08CHU1	08chu1	mus musculu	459	6	1.2	58	16	09WYH6	09wyh6 thermocoga
387	7	1.4	1024	4	014527	014527	homo sapien	460	6	1.2	61	2	09X4G6	09x4g6 vibrio chol
388	7	1.4	1024	5	081577	081577	plasmodium	461	6	1.2	61	10	09M0F1	09m0f1 aradidopsis
389	7	1.4	1046	5	09UDT6	09udt6	homo sapien	462	6	1.2	62	4	096HE3	096he3 homo sapien
390	7	1.4	1046	11	055156	055156	rattus norv	463	6	1.2	63	10	09AXV6	09axv6 brassica na
391	7	1.4	1046	11	0920H8	0920h8	mus musculu	464	6	1.2	63	16	093J37	093j37 streptomyce
392	7	1.4	1047	11	09EP81	09ep81	mus musculu	465	6	1.2	64	2	09ARF6	09arf6 shigella fl
393	7	1.4	1102	5	091VPO	091vpo	aradidopsis	466	6	1.2	64	10	0943H5	0943h5 oryza sativ
394	7	1.4	1132	5	0815C3	0815c3	plasmodium	467	6	1.2	64	16	08G4L2	08g4l2 bifidobacte
395	7	1.4	1164	4	09ULI1	09uli1	homo sapien	468	6	1.2	66	10	09AXV1	09axv1 brassica ol
396	7	1.4	1164	12	089331	089331	zucchini ye	469	6	1.2	66	10	09AXV0	09axv0 brassica ol
397	7	1.4	1164	12	08B6A8	08b6a8	zucchini ye	470	6	1.2	67	10	09AXV2	09axv2 brassica ca
398	7	1.4	1174	5	09XZ6	09xz6	dirosophila	471	6	1.2	67	16	09KRJ4	09krj4 vibrio chol
399	7	1.4	1174	11	08CHA7	08cha7	mus musculu	472	6	1.2	68	6	09BDW4	09bdw4 sus scrofa
400	7	1.4	1219	5	09XZ09	09xz09	dirosophila	473	6	1.2	68	13	090510	090510 fugu rubrip
401	7	1.4	1232	5	08T821	08t821	dictyosteli	474	6	1.2	68	16	08KC22	08kc22 chlorobium
402	7	1.4	1248	16	08U101	08u101	agrobacteri	475	6	1.2	69	2	091RG2	091rg2 synchococc
403	7	1.4	1257	5	08IOH9	08ioh9	dirosophila	476	6	1.2	69	13	08AXL9	08axl9 oncorhynch
404	7	1.4	1274	10	09LV05	09lv05	aradidopsis	477	6	1.2	70	16	08YXQ6	08yxq6 anabaena sp
405	7	1.4	1294	5	08IQ10	08iq10	dirosophila	478	6	1.2	71	17	0917I0	0917i0 synchococc
406	7	1.4	1402	5	09VC63	09vc63	dirosophila	479	6	1.2	71	17	09HMX5	09hmx5 halobacteri
407	7	1.4	1461	16	00S819	00s819	streptomyce	480	6	1.2	73	16	082521	082521 salmonella
408	7	1.4	1553	16	08CK29	08ck29	dirosophila	481	6	1.2	73	17	08TYK6	08tyk6 methanopyru
409	7	1.4	1594	5	09V6W7	09v6w7	dirosophila	482	6	1.2	74	2	0931H1	0931h1 wolinnella s
410	7	1.4	1940	16	09RUC8	09ruc8	deinococcus	483	6	1.2	74	17	08T153	08t153 methanosarc
411	7	1.4	1941	16	08G751	08g751	bifidobacte	484	6	1.2	76	6	08SEP20	08sep20 macaca mula
412	7	1.4	1947	2	09XCJ2	09xcj2	salmonella	485	6	1.2	77	2	08KMH9	08kmh9 mycoplasma
413	7	1.4	2108	13	098U19	098u19	gallus gall	486	6	1.2	77	2	08KHV5	08khv5 mycoplasma
414	7	1.4	2134	12	09YLS4	09yls4	avian encerp	487	6	1.2	77	2	08KH20	08kh20 mycoplasma
415	7	1.4	2275	17	08TSE8	08tse8	methanosarc	488	6	1.2	77	2	08KH21	08kh21 mycoplasma
416	7	1.4	2435	16	08ZNS6	08zns6	salmonella	489	6	1.2	77	2	08KH10	08kh10 mycoplasma
417	7	1.4	2467	4	015050	015050	homo sapien	490	6	1.2	77	11	08R576	08r576 mus musculu
418	7	1.4	2603	6	09X5S3	09x5s3	potorous tr	491	6	1.2	77	16	08X8R3	08x8r3 escherichia
419	7	1.4	2691	4	09NS12	09ns12	homo sapien	492	6	1.2	77	16	08R0W5	08r0w5 corynebacte
420	7	1.4	2692	4	09Y503	09y503	homo sapien	493	6	1.2	77	16	08FL30	08fl30 leptospira
421	7	1.4	2705	4	09S103	09s103	homo sapien	494	6	1.2	77	16	08RANO	08ran0 shewanelia
422	7	1.4	2725	4	09NYES	09nyes	homo sapien	495	6	1.2	77	17	028902	028902 archaeglob
423	7	1.4	3080	12	09WAC2	09wac2	zucchini ye	496	6	1.2	78	15	078638	078638 human immun
424	7	1.4	3413	2	054593	054593	amycolatops	497	6	1.2	78	16	08ZJW7	08zjw7 salmonella
425	7	1.4	4498	5	09W223	09w223	dirosophila	498	6	1.2	78	16	08ZOV3	08zov3 salmonella
426	7	1.4	9477	2	09L4X3	09l4x3	streptomyce	499	6	1.2	79	2	093619	093619 staphylococ
427	7	1.4	11096	2	09L4W3	09l4w3	methanobact	500	6	1.2	79	12	010321	010321 oryza psen
428	6	1.2	17	1	09V2X3	09v2x3	homo sapien	501	6	1.2	79	15	081550	081550 chimpanzee
429	6	1.2	24	4	096D61	096d61	homo sapien	502	6	1.2	80	2	084933	084933 streptococ
430	6	1.2	26	12	09ORV1	09orv1	hepaticis c	503	6	1.2	80	15	08URM2	08urm2 human immun
431	6	1.2	27	13	057557	057557	lampetra pl	504	6	1.2	80	15	08URM3	08urm3 human immun
432	6	1.2	28	2	047354	047354	escherichia	505	6	1.2	80	16	08X8C0	08x8c0 escherichia
433	6	1.2	30	10	08RV13	08rv13	pinus pinas	506	6	1.2	81	4	08U621	08u621 agrobacteri
434	6	1.2	31	11	08CSM5	08csm5	mus musculu	507	6	1.2	81	4	014749	014749 homo sapien
435	6	1.2	34	16	08EBY3	08eby3	shewanelia	508	6	1.2	82	9	08SC60	08sc60 stx2 conver
436	6	1.2	36	16	08DPT5	08dpt5	vibrio vuln	509	6	1.2	82	16	098C08	098c08 rhizobium 1
437	6	1.2	42	2	053518	053518	escherichia	510	6	1.2	82	16	08Z2U2	08z2u2 salmonella
438	6	1.2	42	13	09DF18	09df18	myxococphal	511	6	1.2	83	3	09US96	09us96 schizosacch
439	6	1.2	42	16	092MW2	092mw2	rhizobium m	512	6	1.2	83	5	08WYM9	08wym9 boletenia vi
440	6	1.2	44	2	09FS26	09fs26	escherichia	513	6	1.2	83	16	099U58	099u58 staphylococ
441	6	1.2	47	5	018132	018132	achetina fu	514	6	1.2	84	2	091LAH4	091lah4 thobacillu
442	6	1.2	49	16	08EK96	08ek96	shewanelia	515	6	1.2	84	15	089145	089145 chimpanzee
443	6	1.2	50	10	09AXV4	09axv4	brassica ca	516	6	1.2	84	17	08ZUM5	08zum5 pyrocaculum
444	6	1.2	50	10	09AXV5	09axv5	brassica na	517	6	1.2	84	15	089143	089143 chimpanzee
445	6	1.2	50	16	09PF64	09pf64	xylella fas	518	6	1.2	86	3	096UD1	096ud1 neurospora
446	6	1.2	52	10	09AXV7	09axv7	brassica na	519	6	1.2	87	13	090452	090452 brachydanto
447	6	1.2	53	16	08PJ79	08pj79	xanthomonas	520	6	1.2	87	15	089146	089146 chimpanzee
448	6	1.2	55	4	096F89	096f89	homo sapien	521	6	1.2	88	2	08FXR2	08fxr2 bruceella su
449	6	1.2	55	8	0958E0	0958e0	tinamus maj	522	6	1.2	88	2	09KUR7	09kur7 myxococcus
450	6	1.2	55	10	08S2X3	08s2x3	zea mays (m	523	6	1.2	90	6	0966J2	0966j2 caenorhabdi
451	6	1.2	55	16	09PGK5	09pgk5	xylella fas	524	6	1.2	90	6	002735	002735 sus scrofa
452	6	1.2	56	2	09L7F8	09l7f8	synchococc	525	6	1.2	90	10	091PB1	091pb1 aradidopsis
453	6	1.2	56	10	08RUB4	08rub4	zea mays (m	526	6	1.2	91	7	08SNV9	08snv9 homo sapien
454	6	1.2	57	13	091566	091566	xenopus lae	527	6	1.2	91	7	030205	030205 homo sapien

528	6	1.2	91	7	019515	019515 homo sapien	601	6	1.2	108	16	099XU9	099XJ9 streptococc
529	6	1.2	91	10	09SRX6	09SRX6 arabidopsis	602	6	1.2	108	16	08ZUB9	08ZJES yersinia pe
530	6	1.2	91	16	08F4X0	08F4X0 leprospira	603	6	1.2	108	16	08FCX0	08FCX0 escherichia
531	6	1.2	92	8	09MP63	09MP63 larix smelti	604	6	1.2	109	5	08SV06	08SV06 encephalito
532	6	1.2	92	8	09MP64	09MP64 pseudotsuga	605	6	1.2	109	5	08RTK8	08RTK8 drosophiila
533	6	1.2	92	8	09MP55	09MP55 cycas panzn	606	6	1.2	109	5	08T0J7	08T0J7 drosophiila
534	6	1.2	92	8	09MP66	09MP66 picea smth	607	6	1.2	109	6	09N0W5	09N0W5 oryctolagus
535	6	1.2	92	13	057377	057377 brachydantio	608	6	1.2	109	16	08P564	08P564 xanthomomas
536	6	1.2	92	15	088095	088095 chimpanzee	609	6	1.2	109	17	057992	057992 pyrococcus
537	6	1.2	92	17	025007	025007 methanobact	610	6	1.2	110	5	08T0J7	08T0J7 drosophiila
538	6	1.2	93	16	09PBR0	09PBR0 xyella fas	611	6	1.2	110	16	08ED84	08ED84 shewanella
539	6	1.2	95	4	09URR4	09URR4 homo sapien	612	6	1.2	112	2	09L482	09L482 uncultured
540	6	1.2	95	16	098LX8	098LX8 rhizobium 1	613	6	1.2	112	2	09L481	09L481 uncultured
541	6	1.2	96	4	096117	096117 homo sapien	614	6	1.2	112	2	09L476	09L476 uncultured
542	6	1.2	96	17	08ZAG7	08ZAG7 pyrobaculum	615	6	1.2	112	2	09L479	09L479 uncultured
543	6	1.2	97	2	049073	049073 mycoplasma	616	6	1.2	112	5	08WQJ0	08WQJ0 sporopetra
544	6	1.2	98	11	08R5B8	08R5B8 mus musculu	617	6	1.2	112	5	08TK93	08TK93 drosophiila
545	6	1.2	98	16	08FRK0	08FRK0 escherichia	618	6	1.2	112	16	08ZDR9	08ZDR9 yersinia pe
546	6	1.2	99	2	033447	033447 proteus mir	619	6	1.2	113	6	096934	096934 ceratilis c
547	6	1.2	99	10	08LGD1	08LGD1 arabidopsis	620	6	1.2	113	6	029367	029367 sus scrofa
548	6	1.2	99	12	08OKZ2	08OKZ2 malvastrum	621	6	1.2	113	10	08L4S7	08L4S7 oryza sativ
549	6	1.2	100	5	0963A7	0963A7 shigella fl	622	6	1.2	113	15	09WK54	09WK54 human immun
550	6	1.2	100	5	0963A7	0963A7 leishmania	623	6	1.2	113	15	09WK72	09WK72 human immun
551	6	1.2	100	12	09WB72	09WB72 hepatitis c	624	6	1.2	113	15	09WK52	09WK52 human immun
552	6	1.2	100	12	09WB74	09WB74 hepatitis c	625	6	1.2	113	15	09WK53	09WK53 human immun
553	6	1.2	100	12	09WB56	09WB56 hepatitis c	626	6	1.2	114	2	P71419	P71419 klebsiella
554	6	1.2	100	12	09WB39	09WB39 hepatitis c	627	6	1.2	114	16	09RUX6	09RUX6 deinococcus
555	6	1.2	100	12	09WB70	09WB70 hepatitis c	628	6	1.2	114	17	097BA7	097BA7 thermoplasm
556	6	1.2	100	12	09WB63	09WB63 hepatitis c	629	6	1.2	115	10	08L648	08L648 deschampsia
557	6	1.2	100	12	09WB71	09WB71 hepatitis c	630	6	1.2	115	11	060926	060926 mus musculu
558	6	1.2	100	12	09WB59	09WB59 hepatitis c	631	6	1.2	115	15	056909	056909 human immun
559	6	1.2	100	12	09WB54	09WB54 hepatitis c	632	6	1.2	116	9	09FEV3	09FEV3 bacterioph
560	6	1.2	100	12	09W909	09W909 hepatitis c	633	6	1.2	116	16	09YUAD	09YUAD pseudomonas
561	6	1.2	100	17	09YG98	09YG98 aeropyrum p	634	6	1.2	116	17	09YEM9	09YEM9 aeropyrum p
562	6	1.2	101	11	08BPR9	08BPR9 mus musculu	635	6	1.2	117	2	09R822	09R822 helicobacte
563	6	1.2	101	16	0960E7	0960E7 mycoplasma	636	6	1.2	117	5	027717	027717 plasmodium
564	6	1.2	101	17	09YD64	09YD64 aeropyrum p	637	6	1.2	117	5	08L1V1	08L1V1 plasmodium
565	6	1.2	102	12	059764	059764 schizosacch	638	6	1.2	117	10	0943X9	0943X9 oryza sativ
566	6	1.2	102	12	09J2M0	09J2M0 macaca mula	639	6	1.2	117	15	073225	073225 human immun
567	6	1.2	102	14	08BPN7	08BPN7 uncultured	640	6	1.2	117	15	073223	073223 human immun
568	6	1.2	102	14	08BPN2	08BPN2 uncultured	641	6	1.2	117	16	08DMK1	08DMK1 synchococc
569	6	1.2	102	14	08BPN7	08BPN7 uncultured	642	6	1.2	117	16	08DE56	08DE56 vibrio vuln
570	6	1.2	102	14	08BPM1	08BPM1 uncultured	643	6	1.2	118	9	08HAE8	08HAE8 salmonella
571	6	1.2	102	14	08BFM0	08BFM0 uncultured	644	6	1.2	118	16	09LOF4	09LOF4 streptomyce
572	6	1.2	102	14	08BFL9	08BFL9 uncultured	645	6	1.2	120	10	081056	081056 arabidopsis
573	6	1.2	102	14	08BFL6	08BFL6 uncultured	646	6	1.2	120	16	035601	035601 helicobacte
574	6	1.2	102	14	08BFL3	08BFL3 uncultured	647	6	1.2	120	16	09ZKQ4	09ZKQ4 helicobacte
575	6	1.2	102	14	08BFL1	08BFL1 uncultured	648	6	1.2	120	16	08FRF4	08FRF4 corynebacte
576	6	1.2	102	14	08BFL8	08BFL8 uncultured	649	6	1.2	121	15	09WM41	09WM41 human immun
577	6	1.2	102	14	08BFK7	08BFK7 uncultured	650	6	1.2	121	16	08EVLJ	08EVLJ mycoplasma
578	6	1.2	102	14	08BFL4	08BFL4 uncultured	651	6	1.2	121	17	08EVL0	08EVL0 methanosarc
579	6	1.2	102	14	08BFL3	08BFL3 uncultured	652	6	1.2	122	2	09S0R2	09S0R2 shewanella
580	6	1.2	102	14	08BFL5	08BFL5 uncultured	653	6	1.2	122	2	09REDE	09REDE burkholderi
581	6	1.2	103	10	09XJ71	09XJ71 cucumis sat	654	6	1.2	122	16	08ZLF1	08ZLF1 salmonella
582	6	1.2	103	11	08BOK2	08BOK2 mus musculu	655	6	1.2	122	16	08DEJ8	08DEJ8 vibrio vuln
583	6	1.2	104	6	09SN86	09SN86 ovis aries	656	6	1.2	123	2	045346	045346 bacillus pa
584	6	1.2	104	16	08GCV8	08GCV8 brucella su	657	6	1.2	123	16	09HUX0	09HUX0 pseudomonas
585	6	1.2	105	6	09GMC0	09GMC0 canis famli	658	6	1.2	124	5	09VFO7	09VFO7 drosophiila
586	6	1.2	105	16	08FOI5	08FOI5 leprospira	659	6	1.2	124	16	09RTY4	09RTY4 deinococcus
587	6	1.2	107	2	09APX2	09APX2 pseudomonas	660	6	1.2	124	16	0988A7	0988A7 rhizobium 1
588	6	1.2	107	5	09UZY3	09UZY3 caenorhabdi	661	6	1.2	125	5	08SW51	08SW51 encephalito
589	6	1.2	107	5	08ITL0	08ITL0 drosophiila	662	6	1.2	125	12	08H354	08H354 oryza sativ
590	6	1.2	107	5	08IO53	08IO53 drosophiila	663	6	1.2	125	12	09EME9	09EME9 amastoma moo
591	6	1.2	107	15	09JDX4	09JDX4 human immun	664	6	1.2	125	16	08USK7	08USK7 agrobacteri
592	6	1.2	107	15	09JDX3	09JDX3 human immun	665	6	1.2	125	17	08UL18	08UL18 pyrococcus
593	6	1.2	108	2	09LAI5	09LAI5 thiodiobacilli	666	6	1.2	126	11	09CR06	09CR06 mus musculu
594	6	1.2	108	5	08IO07	08IO07 drosophiila	667	6	1.2	126	12	09WSV2	09WSV2 feline cali
595	6	1.2	108	10	08SSS9	08SSS9 oryza sativ	668	6	1.2	126	15	08URH1	08URH1 human immun
596	6	1.2	108	10	08SSS6	08SSS6 oryza sativ	669	6	1.2	126	15	08URH2	08URH2 human immun
597	6	1.2	108	10	094CY1	094CY1 oryza sativ	670	6	1.2	126	16	09PE51	09PE51 xyella fas
598	6	1.2	108	10	094HY0	094HY0 oryza sativ	671	6	1.2	126	16	09RKS5	09RKS5 bacillus ha
599	6	1.2	108	10	09AWY9	09AWY9 oryza sativ	672	6	1.2	126	16	08YV74	08YV74 anabaena sp
600	6	1.2	108	10	08GSR6	08GSR6 oryza sativ	673	6	1.2	127	4	08NBX4	08NBX4 homo sapien

674	6	1.2	127	10	Q940E4	Q940E4 castanea sa	747	6	1.2	141	16	Q8RFT1	Q8RFT1 fusobacteri
675	6	1.2	127	10	Q22942	Q22942 aradidopsis	748	6	1.2	141	16	Q8KGI9	Q8KGI9 chlorobium
676	6	1.2	127	11	Q70192	Q70192 mus musculu	749	6	1.2	142	2	Q87820	Q87820 rhizobium m
677	6	1.2	127	16	Q97HC2	Q97HC2 clostridium	750	6	1.2	142	2	Q87784	Q87784 ochrobactru
678	6	1.2	127	16	Q8PNO7	Q8PNO7 xanthomonas	751	6	1.2	142	2	Q9AFI3	Q9AFI3 aeromonas h
679	6	1.2	128	16	Q8VK84	Q8VK84 mycobacteri	752	6	1.2	142	2	Q87824	Q87824 rhodobacter
680	6	1.2	129	10	Q9LSH0	Q9LSH0 aradidopsis	753	6	1.2	142	4	Q9NY13	Q9NY13 homo sapien
681	6	1.2	129	10	Q9S7U3	Q9S7U3 aradidopsis	754	6	1.2	142	10	Q8G7B4	Q8G7B4 chlorella v
682	6	1.2	129	10	Q8LDS7	Q8LDS7 aradidopsis	755	6	1.2	142	15	Q9WM55	Q9WM55 human immun
683	6	1.2	129	17	Q9HNS4	Q9HNS4 halobacteri	756	6	1.2	142	15	Q9WM31	Q9WM31 human immun
684	6	1.2	130	2	Q8RK36	Q8RK36 pasteurella	757	6	1.2	142	15	Q9WM62	Q9WM62 human immun
685	6	1.2	130	8	Q8U034	Q8U034 exoneurella	758	6	1.2	142	15	Q9WM53	Q9WM53 human immun
686	6	1.2	130	8	Q8U033	Q8U033 exoneurella	759	6	1.2	142	15	Q9WM48	Q9WM48 human immun
687	6	1.2	130	16	Q9RI71	Q9RI71 streptomyce	760	6	1.2	142	16	Q8YHQ1	Q8YHQ1 bruceella me
688	6	1.2	130	16	Q8EBJ2	Q8EBJ2 shewanella	761	6	1.2	142	16	Q8XJH8	Q8XJH8 clostridium
689	6	1.2	130	16	Q8DW97	Q8DW97 streptococc	762	6	1.2	142	16	Q8XJH8	Q8XJH8 clostridium
690	6	1.2	130	16	Q8CWU4	Q8CWU4 streptococc	763	6	1.2	143	2	P72340	P72340 rhodobacter
691	6	1.2	131	2	P72240	P72240 pseudomonas	764	6	1.2	143	2	Q8KLF3	Q8KLF3 rhizobium e
692	6	1.2	131	5	Q23478	Q23478 caenorhabdi	765	6	1.2	143	9	Q9MCC9	Q9MCC9 lactobacill
693	6	1.2	131	8	Q35865	Q35865 strongyloce	766	6	1.2	143	11	Q9CM84	Q9CM84 mus musculu
694	6	1.2	131	8	Q79125	Q79125 brevineura	767	6	1.2	143	12	Q72518	Q72518 sindbis vir
695	6	1.2	131	10	Q8RZK5	Q8RZK5 oryza sativ	768	6	1.2	143	12	Q72519	Q72519 sindbis vir
696	6	1.2	131	15	Q90823	Q90823 human immun	769	6	1.2	143	12	Q72517	Q72517 sindbis vir
697	6	1.2	132	2	Q85755	Q85755 brochothrix	770	6	1.2	143	12	Q72520	Q72520 sindbis vir
698	6	1.2	132	2	Q8MA75	Q8MA75 buddleja as	771	6	1.2	144	2	Q9E252	Q9E252 brachyrihizob
699	6	1.2	132	10	Q65172	Q65172 mesembryant	772	6	1.2	144	5	Q9GZ13	Q9GZ13 caenorhabdi
700	6	1.2	132	10	Q8G7B3	Q8G7B3 chaetoceros	773	6	1.2	144	5	Q9LENI	Q9LENI cyanophora
701	6	1.2	132	16	Q8XV39	Q8XV39 ralsonia s	774	6	1.2	144	16	Q910G3	Q910G3 pseudomonas
702	6	1.2	132	17	Q8ZXY3	Q8ZXY3 pyrobaculum	775	6	1.2	145	2	Q9RIF1	Q9RIF1 magnetospir
703	6	1.2	133	10	Q8G7B5	Q8G7B5 phaeodactyl	776	6	1.2	145	2	Q8KPI8	Q8KPI8 uncultured
704	6	1.2	133	15	Q9WLU5	Q9WLU5 human immun	777	6	1.2	145	5	Q21324	Q21324 caenorhabdi
705	6	1.2	133	15	Q9WLV7	Q9WLV7 human immun	778	6	1.2	145	16	Q8BN79	Q8BN79 xanthomonas
706	6	1.2	133	15	Q9WLM1	Q9WLM1 human immun	779	6	1.2	146	2	Q9KM02	Q9KM02 pseudomonas
707	6	1.2	133	15	Q9WLU9	Q9WLU9 human immun	780	6	1.2	146	10	Q8LGV9	Q8LGV9 oryza sativ
708	6	1.2	133	15	Q9WLT2	Q9WLT2 human immun	781	6	1.2	146	10	Q9W4W8	Q9W4W8 citrus sine
709	6	1.2	133	15	Q9WLV6	Q9WLV6 human immun	782	6	1.2	146	10	Q9S5X5	Q9S5X5 aradidopsis
710	6	1.2	133	15	Q9YXW1	Q9YXW1 human immun	783	6	1.2	146	10	Q94BY2	Q94BY2 aradidopsis
711	6	1.2	133	15	Q9WLU3	Q9WLU3 human immun	784	6	1.2	146	12	Q80931	Q80931 human papil
712	6	1.2	133	16	Q8XJW7	Q8XJW7 escherichia	785	6	1.2	147	2	Q93R84	Q93R84 vibrio fusc
713	6	1.2	134	2	P71275	P71275 enterobacte	786	6	1.2	147	2	Q8RJ11	Q8RJ11 vibrio loge
714	6	1.2	134	10	Q944N0	Q944N0 olea europ	787	6	1.2	147	2	Q8KSW8	Q8KSW8 streptomyce
715	6	1.2	134	10	Q9AYE3	Q9AYE3 oryza sativ	788	6	1.2	147	2	Q8G7V7	Q8G7V7 citrobacter
716	6	1.2	134	10	Q93XW3	Q93XW3 brassica ca	789	6	1.2	147	6	Q8MKF9	Q8MKF9 sus scrofa
717	6	1.2	134	12	Q9QC03	Q9QC03 simian cyto	790	6	1.2	147	11	Q8K3F3	Q8K3F3 rattus norv
718	6	1.2	134	16	Q93BM1	Q93BM1 streptococc	791	6	1.2	147	16	Q97GZ1	Q97GZ1 anabaena sp
719	6	1.2	134	16	Q8FIAS	Q8FIAS leptospira	792	6	1.2	147	16	Q8YGY4	Q8YGY4 halobacteri
720	6	1.2	135	15	Q9WM86	Q9WM86 human immun	793	6	1.2	147	17	Q9HNNX	Q9HNNX halobacteri
721	6	1.2	135	16	Q8BD51	Q8BD51 vibrio vuln	794	6	1.2	148	2	Q9FAE4	Q9FAE4 alteromonas
722	6	1.2	136	5	Q9VTF3	Q9VTF3 diatophila	795	6	1.2	148	12	Q9DMG5	Q9DMG5 rat cytomeg
723	6	1.2	137	10	Q9SWY3	Q9SWY3 coriandrum	796	6	1.2	148	16	Q92YX3	Q92YX3 rhizobium m
724	6	1.2	137	10	Q8LBM4	Q8LBM4 aradidopsis	797	6	1.2	149	4	Q9BYC9	Q9BYC9 homo sapien
725	6	1.2	137	10	Q9SLE6	Q9SLE6 aradidopsis	798	6	1.2	149	11	Q91Y14	Q91Y14 mus musculu
726	6	1.2	137	15	Q9SESP	Q9SESP human immun	799	6	1.2	149	11	Q9CQL4	Q9CQL4 mus musculu
727	6	1.2	137	16	Q9JMT5	Q9JMT5 escherichia	800	6	1.2	149	11	Q9D5L3	Q9D5L3 mus musculu
728	6	1.2	137	16	Q8ZS54	Q8ZS54 salmonella	801	6	1.2	150	16	Q9KAY3	Q9KAY3 bacillus ha
729	6	1.2	137	16	Q8PIR6	Q8PIR6 streptococc	802	6	1.2	150	16	Q9X872	Q9X872 streptomyce
730	6	1.2	137	16	Q8P0N4	Q8P0N4 streptococc	803	6	1.2	150	16	Q8DAH7	Q8DAH7 vibrio vuln
731	6	1.2	138	4	Q9H7R1	Q9H7R1 homo sapien	804	6	1.2	151	5	Q8ISP3	Q8ISP3 branchiost
732	6	1.2	138	10	Q8HSN8	Q8HSN8 oryza sativ	805	6	1.2	151	13	Q918E5	Q918E5 fugu rubrip
733	6	1.2	138	16	P73673	P73673 synchocyst	806	6	1.2	151	16	Q67294	Q67294 aquilex aeo
734	6	1.2	138	16	Q8E0Z3	Q8E0Z3 streptococc	807	6	1.2	152	4	Q8N124	Q8N124 homo sapien
735	6	1.2	138	16	Q8E0Z3	Q8E0Z3 streptococc	808	6	1.2	152	4	Q8N197	Q8N197 homo sapien
736	6	1.2	139	5	Q49947	Q49947 borrelia bu	809	6	1.2	152	4	Q8NB05	Q8NB05 homo sapien
737	6	1.2	139	10	Q8U652	Q8U652 aradidopsis	810	6	1.2	152	12	Q9PND0	Q9PND0 ct virus. O
738	6	1.2	139	10	Q8G7B2	Q8G7B2 chlorella v	811	6	1.2	152	15	Q8URH4	Q8URH4 human immun
739	6	1.2	139	16	Q8U5I9	Q8U5I9 agrobacteri	812	6	1.2	152	15	Q8URH4	Q8URH4 human immun
740	6	1.2	140	2	Q9RF50	Q9RF50 magnetospir	813	6	1.2	152	15	Q9S5W6	Q9S5W6 streptococc
741	6	1.2	140	10	Q9FMY8	Q9FMY8 aradidopsis	814	6	1.2	152	17	Q97BJ2	Q97BJ2 thermoplaem
742	6	1.2	140	12	Q9OSP5	Q9OSP5 lymphocytic	815	6	1.2	153	2	Q82923	Q82923 escherichia
743	6	1.2	140	16	Q8G4K0	Q8G4K0 bifidobacte	816	6	1.2	153	2	Q930L4	Q930L4 escherichia
744	6	1.2	141	2	Q87687	Q87687 blastobacte	817	6	1.2	153	2	Q8GEP1	Q8GEP1 pseudomonas
745	6	1.2	141	16	Q50907	Q50907 borrelia bu	818	6	1.2	153	10	Q9EFD5	Q9EFD5 aradidopsis
746	6	1.2	141	16	Q9KRVS	Q9KRVS vibrio choi	819	6	1.2	153	11	Q9CXY8	Q9CXY8 mus musculu

820	6	1.2	153	15	Q91WN4	Q91wn4 human immun	893	6	1.2	164	4	Q9BQ52	Q9bq52 homo sapien
821	6	1.2	153	16	Q8ZK03	Q8zk03 salmonella	894	6	1.2	164	5	Q9XUV5	Q9xuv5 caenorhabdi
822	6	1.2	153	16	Q8Z0Y3	Q8z0y3 salmonella	895	6	1.2	164	10	Q8H2Q6	Q8h2q6 oryza sativ
823	6	1.2	154	5	Q8IA35	Q8ia35 brachiolesto	896	6	1.2	165	2	Q934M7	Q934m7 uncultured
824	6	1.2	154	10	Q9S724	Q9s724 arabidopsis	897	6	1.2	165	2	Q934M4	Q934m4 uncultured
825	6	1.2	154	11	Q9D217	Q9d217 mus musculu	898	6	1.2	165	2	Q8RT02	Q8rt02 uncultured
826	6	1.2	154	12	Q9J5S3	Q9j5s3 orangutan h	899	6	1.2	165	2	Q934M6	Q934m6 uncultured
827	6	1.2	154	15	Q9E8G7	Q9e8g7 human immun	900	6	1.2	165	2	Q934M5	Q934m5 uncultured
828	6	1.2	154	15	Q91WN3	Q91wn3 human immun	901	6	1.2	165	5	Q95Q19	Q95q19 caenorhabdi
829	6	1.2	154	15	Q91WN3	Q91wn3 human immun	902	6	1.2	165	6	Q9GLF4	Q9glf4 sus scrofa
830	6	1.2	155	4	Q8WMB8	Q8wmb8 homo sapien	903	6	1.2	165	10	Q65155	Q65155 arabidopsis
831	6	1.2	155	4	Q9HA73	Q9ha73 homo sapien	904	6	1.2	165	16	Q93157	Q93157 staphylococ
832	6	1.2	155	5	Q61523	Q61523 caenorhabdi	905	6	1.2	165	16	Q8ZHJ3	Q8zhj3 yersinia pe
833	6	1.2	155	5	Q8SSF6	Q8ssf6 encephalito	906	6	1.2	165	16	Q8XJD4	Q8xjd4 clostridium
834	6	1.2	155	15	Q91WM1	Q91wm1 human immun	907	6	1.2	165	16	Q8BDY3	Q8bdy3 vibrio vuln
835	6	1.2	155	16	Q92A73	Q92a73 streptomyce	908	6	1.2	165	17	Q9Y8Y9	Q9y8y9 aeropyrum p
836	6	1.2	156	2	Q9ZAT3	Q9zat3 yersinia en	909	6	1.2	166	10	Q9RG16	Q9rg16 arabidopsis
837	6	1.2	156	2	Q9S6G3	Q9s6g3 streptococc	910	6	1.2	166	10	Q8KOM0	Q8kom0 oryza sativ
838	6	1.2	156	2	Q9S6G4	Q9s6g4 streptococc	911	6	1.2	166	16	Q97GF4	Q97gf4 clostridium
839	6	1.2	156	4	Q9JUT8	Q9jut8 homo sapien	912	6	1.2	166	16	Q9X8W3	Q9x8w3 streptomyce
840	6	1.2	156	4	Q9IUX5	Q9iux5 homo sapien	913	6	1.2	166	16	Q9YCD6	Q9ycd6 aeropyrum p
841	6	1.2	156	4	Q9U9U0	Q9u9u0 nereis vire	914	6	1.2	167	2	Q8RLT8	Q8rlt8 neisseria m
842	6	1.2	156	16	Q920F1	Q920f1 rhizobium m	915	6	1.2	167	10	Q8UD06	Q8ud06 arabidopsis
843	6	1.2	156	16	Q8PE65	Q8pe65 xanthomonas	916	6	1.2	167	16	Q8X500	Q8x500 escherichia
844	6	1.2	157	2	Q3J328	Q3j328 acetobacter	917	6	1.2	167	16	Q8CW26	Q8cw26 escherichia
845	6	1.2	157	5	Q81BD9	Q8ied9 plasmodium	918	6	1.2	167	17	Q9HW26	Q9hw26 thermoplas
846	6	1.2	157	5	Q81599	Q81599 plasmodium	919	6	1.2	167	17	Q980H0	Q980h0 sulfolobus
847	6	1.2	157	11	Q9CV79	Q9cv79 mus musculu	920	6	1.2	168	2	Q8KPI9	Q8kpi9 uncultured
848	6	1.2	157	16	Q9L2C7	Q9l2c7 streptomyce	921	6	1.2	168	5	Q8IDG4	Q8idg4 plasmodium
849	6	1.2	157	16	Q8RF73	Q8rf73 corynebacte	922	6	1.2	168	16	Q8KRX5	Q8krx5 chlorobium
850	6	1.2	158	2	Q9S6B7	Q9s6b7 achromobact	923	6	1.2	168	16	Q9S2L1	Q9s2l1 streptomyce
851	6	1.2	158	4	Q9U277	Q9uj27 homo sapien	924	6	1.2	169	2	Q8RT01	Q8rt01 uncultured
852	6	1.2	158	10	Q8GVZ9	Q8gvz9 oryza sativ	925	6	1.2	169	2	Q69016	Q69016 zymomonas m
853	6	1.2	158	16	Q8G1C9	Q8g1c9 brucella su	926	6	1.2	169	16	Q9K3W8	Q9k3w8 streptomyce
854	6	1.2	159	2	Q8G1M5	Q8g1m5 uncultured	927	6	1.2	170	2	Q68295	Q68295 borrelia ja
855	6	1.2	159	2	Q8G1M3	Q8g1m3 uncultured	928	6	1.2	170	10	Q9STH7	Q9stht7 arabidopsis
856	6	1.2	159	2	Q8G1M2	Q8g1m2 uncultured	929	6	1.2	170	16	Q8DM61	Q8dm61 synecococc
857	6	1.2	159	2	Q8G1M1	Q8g1m1 uncultured	930	6	1.2	171	2	Q8KXJ3	Q8kxj3 uncultured
858	6	1.2	159	2	Q8G1M0	Q8g1m0 uncultured	931	6	1.2	171	8	Q8HD83	Q8hdx3 panulirus j
859	6	1.2	159	2	Q8G1L9	Q8g1l9 uncultured	932	6	1.2	171	16	Q9A4Q2	Q9a4q2 caulobacter
860	6	1.2	159	2	Q8GBM1	Q8gbm1 uncultured	933	6	1.2	171	16	Q983C1	Q983c1 rhizobium 1
861	6	1.2	159	2	Q8GBM0	Q8gbm0 uncultured	934	6	1.2	172	2	Q8KXJ8	Q8kxj8 uncultured
862	6	1.2	159	2	Q8GBL9	Q8gbl9 uncultured	935	6	1.2	172	2	Q8KPK2	Q8kpk2 uncultured
863	6	1.2	159	2	Q8GBL7	Q8gbl7 uncultured	936	6	1.2	172	2	Q8KPK5	Q8kpk5 uncultured
864	6	1.2	159	2	Q8GBL4	Q8gbl4 uncultured	937	6	1.2	172	2	Q8KPL8	Q8kpl8 uncultured
865	6	1.2	159	2	Q8GBL5	Q8gbl5 uncultured	938	6	1.2	172	2	Q8KPK8	Q8kpk8 uncultured
866	6	1.2	159	4	Q8G8W6	Q8g8w6 uncultured	939	6	1.2	172	2	Q8KPM4	Q8kpm4 uncultured
867	6	1.2	159	4	Q9UJZ6	Q9ujz6 homo sapien	940	6	1.2	172	2	Q8KPK1	Q8kpk1 uncultured
868	6	1.2	159	4	Q8NF74	Q8nf74 homo sapien	941	6	1.2	172	2	Q8KPK5	Q8kpk5 uncultured
869	6	1.2	159	16	Q9A5V2	Q9a5v2 caulobacter	942	6	1.2	172	2	Q8KPK4	Q8kpk4 uncultured
870	6	1.2	160	2	Q51234	Q51234 neisseria m	943	6	1.2	172	2	Q8KPM0	Q8kpm0 uncultured
871	6	1.2	160	2	Q93RP6	Q93rf6 treponema m	944	6	1.2	172	2	Q8K181	Q8k181 uncultured
872	6	1.2	160	2	Q51235	Q51235 neisseria m	945	6	1.2	172	2	Q8KPM5	Q8kpm5 uncultured
873	6	1.2	160	2	Q8GN70	Q8gn70 bifidobacte	946	6	1.2	172	2	Q8KPK9	Q8kpk9 uncultured
874	6	1.2	160	11	Q8C1H8	Q8c1h8 mus musculu	947	6	1.2	172	2	Q8KPL6	Q8kpl6 uncultured
875	6	1.2	160	16	Q9X7R8	Q9x7r8 streptomyce	948	6	1.2	172	2	Q8KPM0	Q8kpm0 uncultured
876	6	1.2	161	4	Q9UK20	Q9uk20 homo sapien	949	6	1.2	172	2	Q8KPL9	Q8kpl9 uncultured
877	6	1.2	161	5	Q8SSP6	Q8ssp6 dictyosteli	950	6	1.2	172	2	Q8KPL2	Q8kpl2 uncultured
878	6	1.2	161	16	Q98CA9	Q98ca9 rhizobium 1	951	6	1.2	172	2	Q8KPL3	Q8kpl3 uncultured
879	6	1.2	161	16	Q92RM5	Q92rms rhizobium m	952	6	1.2	172	2	Q8KPK6	Q8kpk6 uncultured
880	6	1.2	161	16	Q8KRX2	Q8kx2 chlorobium	953	6	1.2	172	2	Q8KPM9	Q8kpm9 uncultured
881	6	1.2	162	5	Q8SWK4	Q8swk4 encephalito	954	6	1.2	172	2	Q8KPL4	Q8kpl4 uncultured
882	6	1.2	162	10	Q8S6B2	Q8s6b2 oryza sativ	955	6	1.2	172	2	Q8KPL5	Q8kpl5 uncultured
883	6	1.2	162	11	Q9CX67	Q9cx67 mus musculu	956	6	1.2	172	2	Q8KPM3	Q8kpm3 uncultured
884	6	1.2	162	16	Q9CJ98	Q9cj98 lactococcus	957	6	1.2	172	2	Q8KPM8	Q8kpm8 uncultured
885	6	1.2	162	16	Q8PEC5	Q8pec5 xanthomonas	958	6	1.2	172	2	Q9R9E8	Q9r9e8 pseudomonas
886	6	1.2	162	16	Q8EAB0	Q8eab0 shewanella	959	6	1.2	172	2	Q8K180	Q8k180 uncultured
887	6	1.2	163	17	Q8PRG4	Q8prg4 methanosarc	960	6	1.2	172	2	Q8KXJ9	Q8kxj9 uncultured
888	6	1.2	163	4	Q8NNH9	Q8nnh9 homo sapien	961	6	1.2	172	2	Q8KPM1	Q8kpm1 uncultured
889	6	1.2	163	10	P93330	P93330 medicago tr	962	6	1.2	172	2	Q8KXJ4	Q8kxj4 uncultured
890	6	1.2	163	10	O64977	O64977 arabidopsis	963	6	1.2	172	2	Q8KPM2	Q8kpm2 uncultured
891	6	1.2	163	17	O9YAY7	O9yay7 aeropyrum p	964	6	1.2	172	2	Q8KPK7	Q8kpk7 uncultured
892	6	1.2	164	2	Q8RTQ9	Q8rtq9 pseudomonas	965	6	1.2	172	5	Q9T228	Q9t228 caenorhabdi

966	6	1.2	172	10	Q8S1B6	Q8S1B6 oryza sativ
967	6	1.2	172	15	P89944	P89944 rabbit endo
968	6	1.2	172	16	Q8ZDL0	Q8ZDL0 yersinia pe
969	6	1.2	172	16	Q8CZ88	Q8CZ88 streptococ
970	6	1.2	173	2	Q8RT05	Q8RT05 uncultured
971	6	1.2	173	2	Q8RT06	Q8RT06 uncultured
972	6	1.2	173	2	Q8RT04	Q8RT04 uncultured
973	6	1.2	173	2	Q8RT03	Q8RT03 uncultured
974	6	1.2	173	12	Q8RT07	Q8RT07 uncultured
975	6	1.2	173	12	Q8RT05	Q8RT05 oat golden
976	6	1.2	173	16	Q8RCX6	Q8RCX6 thermoaer
977	6	1.2	173	16	Q8G4M9	Q8G4M9 bifidobacte
978	6	1.2	174	15	Q8QDK4	Q8QDK4 chimpanzee
979	6	1.2	174	15	Q8QDK3	Q8QDK3 chimpanzee
980	6	1.2	174	15	Q8QDK1	Q8QDK1 chimpanzee
981	6	1.2	174	15	Q8QDK6	Q8QDK6 chimpanzee
982	6	1.2	174	15	Q8QDK6	Q8QDK6 chimpanzee
983	6	1.2	174	17	Q8PDL6	Q8PDL6 escherichia
984	6	1.2	175	9	Q03942	Q03942 halobacteri
985	6	1.2	175	10	Q03V27	Q03V27 zea mays su
986	6	1.2	175	10	Q03W69	Q03W69 zea mays su
987	6	1.2	175	16	Q9PNW0	Q9PNW0 campylobact
988	6	1.2	175	16	Q8ZGS3	Q8ZGS3 yersinia pe
989	6	1.2	176	16	Q9Z224	Q9Z224 rhizobium m
990	6	1.2	176	16	Q99RV1	Q99RV1 staphylococ
991	6	1.2	176	16	Q8CRM3	Q8CRM3 yersinia pe
992	6	1.2	177	2	Q9X690	Q9X690 mercury res
993	6	1.2	177	6	Q9GLF8	Q9GLF8 trichosurus
994	6	1.2	177	8	Q8ML25	Q8ML25 hesperantha
995	6	1.2	177	17	Q9Y911	Q9Y911 aeropyrum p
996	6	1.2	178	2	Q8RLS9	Q8RLS9 chrobacteru
997	6	1.2	178	9	Q9F257	Q9F257 bacterioph
998	6	1.2	178	10	Q94LE8	Q94LE8 oryza sativ
999	6	1.2	178	16	Q9RT80	Q9RT80 deinococcus
1000	6	1.2	178	16	Q8XNH4	Q8XNH4 clostridium

## ALIGNMENTS.

## RESULT 1

Q9UTB8 PRELIMINARY; PRT; 386 AA.  
 AC Q9UTB8;  
 DT 01-OCT-2000 (TREMblrel. 15, Created)  
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)  
 DT 01-OCT-2002 (TREMblrel. 22, Last annotation update)  
 DE Nitrite reductase, major outer membrane copper-containing protein  
 DE (EC 1.7.99.3).  
 GN ANIA OR NMA1887.  
 OS Neisseria meningitidis (serogroup A).  
 OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;  
 OC Neisseriaceae; Neisseria.  
 OX NCBI\_TaxID=65699;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=22491 / Serogroup A / Serotype 4A;  
 RX MEDLINE=2022556; PubMed=10761919;  
 RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,  
 RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,  
 RA Davies R.M., Davis P., Devlin K., Felwell T., Hamlin N., Holtroyd S.,  
 RA Jørgensen K., Leather S., Moule S., Mungall K., Quail M.A.,  
 RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,  
 RA Whitehead S., Spratt B.G., Barrall B.G.;  
 RT "Complete DNA sequence of a serogroup A strain of Neisseria  
 meningitidis Z2491.";  
 RL Nature 404:502-506(2000).  
 DR EMBL; AL162757; CAB85110.1; -;  
 DR HSSP; P38501; 1AS8.  
 DR InterPro; IPR001117; Cu-oxidase.  
 DR InterPro; IPR001287; CUNO2\_reductase.  
 DR Pfam; PF00394; Cu-oxidase; 2.  
 DR PRINTS; PR00695; CUNO2RDTASE.

KW Oxidoreductase; Complete proteome.  
 SQ SEQUENCE 386 AA; 40389 MW; D0624ED3A3979E88 CRC64;  
 Query Match 6.8%; Score 34; DB 16; Length 386;  
 Best Local Similarity 100.0%; Pred. No. 1e-25;  
 Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

## RESULT 2

Q9UTB8 PRELIMINARY; PRT; 390 AA.  
 AC Q9UTB8;  
 DT 01-OCT-2000 (TREMblrel. 15, Created)  
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)  
 DT 01-OCT-2002 (TREMblrel. 22, Last annotation update)  
 DE Major anaerobically induced outer membrane protein.  
 GN NMB1623.  
 OS Neisseria meningitidis (serogroup B).  
 OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;  
 OC Neisseriaceae; Neisseria.  
 OX NCBI\_TaxID=491;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MC58 / Serogroup B;  
 RX MEDLINE=20175755; PubMed=10710307;  
 RA Tettein H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,  
 RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,  
 RA Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,  
 RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,  
 RA Mason T., Cleckler A., Parksey D.S., Blair E., Citlone H., Clark E.B.,  
 RA Cotton M.D., Uitterlbeck T.R., Khouri H., Qin H., Vamathevan J.,  
 RA Gill J., Scarlata V., Maignan V., Pappa M., Grandi G., Sun L.,  
 RA Smith H.O., Fraser C.M., Moxon E.R., Rappelli R., Venter J.C.;  
 RT "Complete genome sequence of Neisseria meningitidis serogroup B strain  
 MC58.";  
 RL Science 287:1809-1815(2000).  
 DR EMBL; AEO02512; AAP41975.1; -;  
 DR HSSP; P25006; INTF.  
 DR TIGR; NMB1623; -;  
 DR InterPro; IPR001117; Cu-oxidase.  
 DR InterPro; IPR001287; CUNO2\_reductase.  
 DR Pfam; PF00394; Cu-oxidase; 2.  
 DR PRINTS; PR00695; CUNO2RDTASE.  
 KW Complete proteome.  
 SQ SEQUENCE 390 AA; 40763 MW; C503F9D47DD1169D CRC64;  
 Query Match 6.8%; Score 34; DB 16; Length 390;  
 Best Local Similarity 100.0%; Pred. No. 1e-25;  
 Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 178 YHCAPVPGMHIANGYGLIVEPKGLPKVDKE 211  
 DB 173 YHCAPVPGMHIANGYGLIVEPKGLPKVDKE 206

## RESULT 3

Q8XRY3 PRELIMINARY; PRT; 510 AA.  
 AC Q8XRY3;  
 DT 01-MAR-2002 (TREMblrel. 20, Created)  
 DT 01-MAR-2002 (TREMblrel. 20, Last sequence update)  
 DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)  
 DE Probable major ANAEROBICALLY induced outer membrane transmembrane  
 DE protein (EC 1.7.99.3).  
 GN RSP1503 OR RSO3038.  
 OS Ralstonia solanacearum (Pseudomonas solanacearum).  
 OG Bacteroid megaplasmid.  
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
 OC Ralstoniaceae; Ralstonia.

```

OX NCBI_TaxID=305;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=GM11000;
RX MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Arlet M., Billault A., Brotier P., Camus J.C., Catolico L.,
RA Chandler M., Choise N., Claudel-Renard C., Cunac S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T.,
RA Siguer P., Thebaud P., Whalen M., Wincker P., Levy M.,
RA Weisenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL Nature 415:497-502(2002).
DR EMBL; AL646085; CAD18654.1; -.
DR InterPro; IPR001117; Cu-oxidase.
DR InterPro; IPR001287; CuNO2 reductase.
DR InterPro; IPR000345; CytC heme_bind.
DR InterPro; IPR003088; Cyt_C1.
DR Pfam; PF00394; Cu-oxidase; 1.
DR Pfam; PF00034; Cytochrome_c; 1.
DR PRINTS; PR00695; CUNO2RDTASE.
DR PROSITE; PS00190; CYTOCHROME_C; 1.
KW Oxidoreductase; Plasmid; Complete proteome.
SQ SEQUENCE 510 AA; 54600 MW; 8008105DD99459AC CRC64;

Query Match 3.8%; Score 19; DB 16; Length 510;
Best Local Similarity 100.0%; Pred. No. 2.9e-10;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 183 APVGMHANGMYGLIVER 201
DB 180 APVGMHANGMYGLIVER 198
|||||
RESULT 4
Q8GBL8 PRELIMINARY; PRT; 159 AA.
ID Q8GBL8;
AC Q8GBL8;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Copper-containing nitrite reductase (Fragment).
GN NIRK.
OS uncultured bacterium.
OC Bacteria; environmental samples.
OX NCBI_TaxID=77133;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=22293558; PubMed=12406765;
RA Avrahami S., Conrad R., Braker G.;
RT "Effect of Soil Ammonium Concentration on N(2)O Release and on the
RT Community Structure of Ammonia Oxidizers and Denitrifiers.";
RL Appl. Environ. Microbiol. 68:5685-5692(2002).
DR EMBL; AJ487549; CAD31803.1; -.
FT NON_TER 1 1
FT NON_TER 159 159
SQ SEQUENCE 159 AA; 17444 MW; 560E25D46ED38B4 CRC64;

Query Match 3.2%; Score 16; DB 2; Length 159;
Best Local Similarity 100.0%; Pred. No. 1.1e-07;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 252 VGALTGENALKAKVE 267
DB 79 VGALTGENALKAKVE 94
|||||
RESULT 5
Q8BFN8 PRELIMINARY; PRT; 94 AA.
ID Q8BFN8;
AC Q8BFN8;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)

```

```

DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Nitrite reductase (Fragment).
GN NIRK.
OS uncultured organism.
OC unclassified; environmental samples.
OX NCBI_TaxID=155900;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=A02-03-168;
RA Fields M.W., Yan T., Wu L., Zu Y., Tiedje J.M., Zhou J.;
RT "Molecular diversity and characterization of nitrite reductase gene
RT fragments (nirK and nirS) from nitrate and uranium-contaminated
RT groundwater.";
RL Environ. Microbiol. 0:0-0(2002).
DR EMBL; AF548914; AAN59838.1; -.
FT NON_TER 1 1
FT NON_TER 94 94
SQ SEQUENCE 94 AA; 10418 MW; A79EDF8DD74B444 CRC64;

Query Match 2.4%; Score 12; DB 14; Length 94;
Best Local Similarity 100.0%; Pred. No. 0.00086;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 252 VGALTGENALKA 263
DB 55 VGALTGENALKA 66
|||||
RESULT 6
Q8BFP2 PRELIMINARY; PRT; 95 AA.
ID Q8BFP2;
AC Q8BFP2;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Nitrite reductase (Fragment).
GN NIRK.
OS uncultured organism.
OC unclassified; environmental samples.
OX NCBI_TaxID=155900;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=A07-16-44;
RA Fields M.W., Yan T., Wu L., Zu Y., Tiedje J.M., Zhou J.;
RT "Molecular diversity and characterization of nitrite reductase gene
RT fragments (nirK and nirS) from nitrate and uranium-contaminated
RT groundwater.";
RL Environ. Microbiol. 0:0-0(2002).
DR EMBL; AF548909; AAN59833.1; -.
FT NON_TER 1 1
FT NON_TER 95 95
SQ SEQUENCE 95 AA; 10559 MW; CF4B42852051EC74 CRC64;

Query Match 2.4%; Score 12; DB 14; Length 95;
Best Local Similarity 100.0%; Pred. No. 0.00087;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 252 VGALTGENALKA 263
DB 56 VGALTGENALKA 67
|||||
RESULT 7
Q8BFL7 PRELIMINARY; PRT; 99 AA.
ID Q8BFL7;
AC Q8BFL7;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Nitrite reductase (Fragment).
GN NIRK.
OS uncultured organism.
OC unclassified; environmental samples.

```

```

OK NCBI_TaxID=155900;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=G01-03-134;
RA Fields M.W., Yan T., Wu L., Zu Y., Tiedje J.M., Zhou J.;
RT "Molecular diversity and characterization of nitrile reductase gene
   fragments (nitrk and nitrS) from nitrate and uranium-contaminated
   groundwater.";
RL Environ. Microbiol. 0:0-0(2002).
DR EMBL; AF548937; AAN59861.1; -.
FT NON_TER 1
FT NON_TER 99
SQ SEQUENCE 99 AA; 11019 MW; AA64C0FC7CB74CA9 CRC64;

Query Match 2.4%; Score 12; DB 14; Length 99;
Best Local Similarity 100.0%; Pred. No. 0.00091;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 252 VGALTGENALKA 263
DB 60 VGALTGENALKA 71

RESULT 8
O8BFN9 PRELIMINARY; PRT; 102 AA.
ID O8BFN9
AC O8BFN9;
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Nitrite reductase (Fragment).
GN NITRK.
OS unclassified organism.
OC unclassified; environmental samples.
OX NCBI_TaxID=155900;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A03-10-5;
RA Fields M.W., Yan T., Wu L., Zu Y., Tiedje J.M., Zhou J.;
RT "Molecular diversity and characterization of nitrile reductase gene
   fragments (nitrk and nitrS) from nitrate and uranium-contaminated
   groundwater.";
RL Environ. Microbiol. 0:0-0(2002).
DR EMBL; AF548913; AAN59837.1; -.
FT NON_TER 1
FT NON_TER 102
SQ SEQUENCE 102 AA; 11327 MW; 9P317491059A3E67 CRC64;

Query Match 2.4%; Score 12; DB 14; Length 102;
Best Local Similarity 100.0%; Pred. No. 0.00093;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 252 VGALTGENALKA 263
DB 63 VGALTGENALKA 74

RESULT 9
O8BFN9 PRELIMINARY; PRT; 102 AA.
ID O8BFN9
AC O8BFN9;
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Nitrite reductase (Fragment).
GN NITRK.
OS unclassified organism.
OC unclassified; environmental samples.
OX NCBI_TaxID=155900;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C01-03-6;
RA Fields M.W., Yan T., Wu L., Zu Y., Tiedje J.M., Zhou J.;

```

```

RT      "Molecular diversity and characterization of nitrite reductase gene
RT      fragments (nirk and nirs) from nitrate and uranium-contaminated
RT      groundwater".;
DR      Environ. Microbiol. 0:0-0(2002).
RL      EMBL; AF548917; AAN59841.1; -.
FT      NON TER
FT      1
FT      102
SQ      SEQUENCE 102 AA; 11256 MW; 090DD6931876E087 CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 0.00093;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      252 VGALTGENALKA 263
      |||||
      63 VGALTGENALKA 74

RESULT 10
Q8BFN3 PRELIMINARY; PRT; 102 AA.
AC Q8BFN3;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Nitrite reductase (Fragment).
GN NIRK.
OS unclutred organism.
OC unclassified; environmental samples.
OX NCI_TaxID=155900;

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=B06-16-18;
RA Fields M.W., Yan T., Wu L., Zu Y., Tiedje J.M., Zhou J.;
RT "Molecular diversity and characterization of nitrite reductase gene
RT fragments (nirk and nirs) from nitrate and uranium-contaminated
RT groundwater".;
RL Environ. Microbiol. 0:0-0(2002).
RL EMBL; AF548919; AAN59843.1; -.
FT NON TER
FT 1
FT 102
SQ SEQUENCE 102 AA; 11245 MW; 915AB0E41975C548 CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 0.00093;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      252 VGALTGENALKA 263
      |||||
      63 VGALTGENALKA 74

RESULT 11
Q8BFNO PRELIMINARY; PRT; 102 AA.
AC Q8BFNO;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Nitrite reductase (Fragment).
GN NNRK.
OS unclutred organism.
OC unclassified; environmental samples.
OX NCI_TaxID=155900;

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=B02-05-26;
RA Fields M.W., Yan T., Wu L., Zu Y., Tiedje J.M., Zhou J.;
RT "Molecular diversity and characterization of nitrite reductase gene
RT fragments (nirk and nirs) from nitrate and uranium-contaminated
RT groundwater".;
RL Environ. Microbiol. 0:0-0(2002).
RL EMBL; AF548922; AAN59846.1; -.

```



FT NON\_TER 1 1  
FT NON\_TER 102 102  
SQ SEQUENCE 102 AA; 11276 MW; 333DF6992990FAB4 CRC64;

Query Match 2.4%; Score 12; DB 14; Length 102;  
Best Local Similarity 100.0%; Pred. No. 0.00093;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 252 VGALTGENALKA 263  
DB 63 VGALTGENALKA 74

## RESULT 12

Q8BFM8 ID O8BFM8 PRELIMINARY; PRT; 102 AA.  
AC O8BFM8;  
DT 01-MAR-2003 (TRENBLREL. 23, Created)  
DT 01-MAR-2003 (TRENBLREL. 23, Last sequence update)  
DE Nitrite reductase (Fragment).  
GN NIRK.  
OS uncultured organism.  
OC unclassified; environmental samples.  
OX NCB1\_TaxID=155900;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=D02-05-81;  
RA Fields M.W., Yan T., Wu L., Zu Y., Tiedje J.M., Zhou J.;  
RT "Molecular diversity and characterization of nitrite reductase gene fragments (nirK and nirS) from nitrate and uranium-contaminated groundwater.";  
RT Environ. Microbiol. 0:0-0(2002).  
DR EMBL; AFS48925; AAN59849.1; -.  
FT NON\_TER 1 1  
FT NON\_TER 102 102  
SQ SEQUENCE 102 AA; 11286 MW; DFD696551IDCC825 CRC64;

Query Match 2.4%; Score 12; DB 14; Length 102;  
Best Local Similarity 100.0%; Pred. No. 0.00093;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 252 VGALTGENALKA 263  
DB 63 VGALTGENALKA 74

## RESULT 13

Q8BFM2 ID O8BFM2 PRELIMINARY; PRT; 102 AA.  
AC O8BFM2;  
DT 01-MAR-2003 (TRENBLREL. 23, Created)  
DT 01-MAR-2003 (TRENBLREL. 23, Last sequence update)  
DT 01-MAR-2003 (TRENBLREL. 23, Last annotation update)  
DE Nitrite reductase (Fragment).  
GN NIRK.  
OS uncultured organism.  
OC unclassified; environmental samples.  
OX NCB1\_TaxID=155900;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=E01-03-40;  
RA Fields M.W., Yan T., Wu L., Zu Y., Tiedje J.M., Zhou J.;  
RT "Molecular diversity and characterization of nitrite reductase gene fragments (nirK and nirS) from nitrate and uranium-contaminated groundwater.";  
RT Environ. Microbiol. 0:0-0(2002).  
DR EMBL; AFS48932; AAN59856.1; -.  
FT NON\_TER 1 1  
FT NON\_TER 102 102  
SQ SEQUENCE 102 AA; 11254 MW; 01AF71COPDD04069 CRC64;

Query Match 2.4%; Score 12; DB 14; Length 102;

Best Local Similarity 100.0%; Pred. No. 0.00093;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 252 VGALTGENALKA 263  
DB 63 VGALTGENALKA 74

## RESULT 14

Q8BFM4 ID O8BFM4 PRELIMINARY; PRT; 102 AA.  
AC O8BFM4;  
DT 01-MAR-2003 (TRENBLREL. 23, Created)  
DT 01-MAR-2003 (TRENBLREL. 23, Last sequence update)  
DT 01-MAR-2003 (TRENBLREL. 23, Last annotation update)  
DE Nitrite reductase (Fragment).  
GN NIRK.  
OS uncultured organism.  
OC unclassified; environmental samples.  
OX NCB1\_TaxID=155900;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=F02-05-104;  
RA Fields M.W., Yan T., Wu L., Zu Y., Tiedje J.M., Zhou J.;  
RT "Molecular diversity and characterization of nitrite reductase gene fragments (nirK and nirS) from nitrate and uranium-contaminated groundwater.";  
RT Environ. Microbiol. 0:0-0(2002).  
DR EMBL; AFS48940; AAN59864.1; -.  
FT NON\_TER 1 1  
FT NON\_TER 102 102  
SQ SEQUENCE 102 AA; 11304 MW; 7972A3661796DE88 CRC64;

Query Match 2.4%; Score 12; DB 14; Length 102;  
Best Local Similarity 100.0%; Pred. No. 0.00093;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 252 VGALTGENALKA 263  
DB 63 VGALTGENALKA 74

## RESULT 15

Q8BFK9 ID O8BFK9 PRELIMINARY; PRT; 102 AA.  
AC O8BFK9;  
DT 01-MAR-2003 (TRENBLREL. 23, Created)  
DT 01-MAR-2003 (TRENBLREL. 23, Last sequence update)  
DT 01-MAR-2003 (TRENBLREL. 23, Last annotation update)  
DE Nitrite reductase (Fragment).  
GN NIRK.  
OS uncultured organism.  
OC unclassified; environmental samples.  
OX NCB1\_TaxID=155900;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=H01-03-137;  
RA Fields M.W., Yan T., Wu L., Zu Y., Tiedje J.M., Zhou J.;  
RT "Molecular diversity and characterization of nitrite reductase gene fragments (nirK and nirS) from nitrate and uranium-contaminated groundwater.";  
RT Environ. Microbiol. 0:0-0(2002).  
DR EMBL; AFS48946; AAN59869.1; -.  
FT NON\_TER 1 1  
FT NON\_TER 102 102  
SQ SEQUENCE 102 AA; 11334 MW; 091839693CB90AB7 CRC64;

Query Match 2.4%; Score 12; DB 14; Length 102;  
Best Local Similarity 100.0%; Pred. No. 0.00093;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 252 VGALTGENALKA 263

Db 63 VgALTGENALKA 74

## RESULT 16

Q8BF60 PRELIMINARY; PRT; 102 AA.  
AC Q8BF60;  
DT 01-MAR-2003 (TEMBLrel. 23, Created)  
DT 01-MAR-2003 (TEMBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)  
DE Nitrile reductase (Fragment).  
GN NIRK.  
OS uncultured organism.  
OC unclassified; environmental samples.  
OX NCBI\_TaxID=155900;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=D03-10-15, E03-10-26, G03-10-86, K16-129, and K05-27;  
RA Fields M.W., Yan T., Wu L., Zu Y., Tiedje J.M., Zhou J.;  
RT "Molecular diversity and characterization of nitrile reductase gene fragments (nirK and nirS) from nitrate and uranium-contaminated groundwater";  
RL Environ. Microbiol. 0:0-0(2002).  
DR EMBL; AF548924; AAN5948.1; -;  
DR EMBL; AF548931; AAN5955.1; -;  
DR EMBL; AF548949; AAN59872.1; -;  
DR EMBL; AF548953; AAN59876.1; -;  
DR EMBL; AF548956; AAN59879.1; -;  
FT NON\_TER 1 1  
FT 102 102  
SQ SEQUENCE 102 AA; 11286 MW; 090DF6831976E087 CRC64;

Query Match 2.4%; Score 12; DB 14; Length 102;  
Best Local Similarity 100.0%; Pred. No. 0.00093;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 252 VgALTGENALKA 263  
Db 63 VgALTGENALKA 74

RESULT 17  
Q8BFP0 PRELIMINARY; PRT; 103 AA.  
AC Q8BFP0;  
DT 01-MAR-2003 (TEMBLrel. 23, Created)  
DT 01-MAR-2003 (TEMBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)  
DE Nitrile reductase (Fragment).  
GN NIRK.  
OS uncultured organism.  
OC unclassified; environmental samples.  
OX NCBI\_TaxID=155900;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A04-15-9;  
RA Fields M.W., Yan T., Wu L., Zu Y., Tiedje J.M., Zhou J.;  
RT "Molecular diversity and characterization of nitrile reductase gene fragments (nirK and nirS) from nitrate and uranium-contaminated groundwater";  
RL Environ. Microbiol. 0:0-0(2002).  
DR EMBL; AF548912; AAN59836.1; -;  
FT NON\_TER 1 1  
FT 103 103  
SQ SEQUENCE 103 AA; 11327 MW; 40EA68FAC498F3E CRC64;

Query Match 2.4%; Score 12; DB 14; Length 103;  
Best Local Similarity 100.0%; Pred. No. 0.00094;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 252 VgALTGENALKA 263  
Db 64 VgALTGENALKA 75

## RESULT 18

O87751 PRELIMINARY; PRT; 142 AA.  
AC O87751;  
DT 01-NOV-1998 (TEMBLrel. 08, Created)  
DT 01-NOV-1998 (TEMBLrel. 08, Last sequence update)  
DT 01-JUN-2002 (TEMBLrel. 21, Last annotation update)  
DE Nitrile reductase (Fragment).  
GN NIRK.  
OS Hyphomicrobium zavarzini.  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
OC Hyphomicrobiaceae; Hyphomicrobium.  
OX NCBI\_TaxID=48292;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=IPAM ZV-622;  
RA MEDLINE=98432797; PubMed=9758798;  
RA Braker G., Fesefeldt A., Witzel K.P.;  
RT "Development of PCR primer systems for the amplification of nitrile reductase genes (nirS and nirK) to detect denitrifying bacteria in environmental samples";  
RL Appl. Environ. Microbiol. 64:3769-3775(1998).  
DR EMBL; AJ224902; CAA12205.1; -;  
DR HSSP; P25006; INIF.  
DR InterPro; IPR001117; Cu-oxidase.  
DR Pfam; PF00394; Cu-oxidase; 1.  
FT NON\_TER 1 1  
FT 142 142  
SQ SEQUENCE 142 AA; 15583 MW; F76814D01D4FD1BA CRC64;

Query Match 2.4%; Score 12; DB 2; Length 142;  
Best Local Similarity 100.0%; Pred. No. 0.0013;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 252 VgALTGENALKA 263  
Db 72 VgALTGENALKA 83

RESULT 19  
Q8KPL1 PRELIMINARY; PRT; 172 AA.  
AC Q8KPL1;  
DT 01-OCT-2002 (TEMBLrel. 22, Created)  
DT 01-OCT-2002 (TEMBLrel. 22, Last sequence update)  
DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)  
DE Nitrile reductase (Fragment).  
GN NIRK.  
OS uncultured bacterium.  
OC Bacteria; environmental samples.  
OX NCBI\_TaxID=77133;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC MEDLINE=21914100; PubMed=11916709;  
RA Prieme A., Braker G., Tiedje J.M.;  
RT "Diversity of nitrile reductase (nirK and nirS) gene fragments in forested upland and wetland soils";  
RL Appl. Environ. Microbiol. 68:1893-1900(2002).  
DR EMBL; AY121543; AAM94074.1; -;  
DR InterPro; IPR001117; Cu-oxidase.  
DR Pfam; PF00394; Cu-oxidase; 1.  
FT NON\_TER 1 1  
FT 172 172  
SQ SEQUENCE 172 AA; 18832 MW; 8A7D4E22E814B686 CRC64;

Query Match 2.4%; Score 12; DB 2; Length 172;  
Best Local Similarity 100.0%; Pred. No. 0.0015;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 252 VgALTGENALKA 263  
Db 72 VgALTGENALKA 83

Db 85 VGALTGENALKA 96

## RESULT 20

Q9A007 PRELIMINARY; PRT; 360 AA.  
 AC Q9A007  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Dissimilatory nitrite reductase.  
 GN NIRK.  
 OS Alcaligenes sp. STC1.  
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
 OC Alcaligenaceae; Alcaligenes.  
 OX NCBI\_TaxID=133923;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=STC1;  
 RA Shoun H., Takaya N.;  
 RT "Alcaligenes sp. STC1 nitrite reductase gene";  
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AB046603; BAB21510.1; -.  
 DR HSSP: P25006; INIF.  
 DR InterPro: IPR001117; Cu-oxidase.  
 DR InterPro: IPR001287; CuNO2\_reductase.  
 DR Pfam: PF00394; Cu-oxidase; 2.  
 DR PRINTS: PR00695; CUNO2RDTASE.  
 DR PROSITE: PS00283; SOYBEAN\_KUNITZ; 1.  
 SQ SEQUENCE 360 AA; 39072 MW; CAF063CB38430380 CRC64;

Query Match 2.4%; Score 12; DB 2; Length 360;  
 Best Local Similarity 100.0%; Pred. No. 0.003;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 137 MPNVDFFHAATG 148  
 Db 111 MPNVDFFHAATG 122

## RESULT 21

Q8KKH4 PRELIMINARY; PRT; 486 AA.  
 AC Q8KKH4  
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Nitrite reductase precursor.  
 GN NIRK.  
 OS Hyphomicrobium denitrificans.  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 OC Hyphomicrobiaceae; Hyphomicrobium.  
 OX NCBI\_TaxID=53399;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=AI151;  
 RA Katakura K., Fukui A., Kabayashi M., Yamaguchi K., Suzuki S.;  
 RT "Cloning and expression of copper-containing nitrite reductase from  
 RT Hyphomicrobium denitrificans."  
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AB076606; BAC00912.1; -.  
 DR InterPro: IPR001117; Cu-oxidase.  
 DR InterPro: IPR001287; CuNO2\_reductase.  
 DR PRINTS: PR00695; CUNO2RDTASE.  
 DR PROSITE: PS00079; MULTICOPPER\_OXIDASE1; 1.  
 KW Signal.  
 FT SIGNAL.  
 SQ SEQUENCE 486 AA; 52238 MW; 6984E02ECF791AF CRC64;

Query Match 2.4%; Score 12; DB 2; Length 486;  
 Best Local Similarity 100.0%; Pred. No. 0.004;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 281 TSSFHVIGEIFD 292  
 Db 400 TSSFHVIGEIFD 411

## RESULT 22

Q8BFL0 PRELIMINARY; PRT; 102 AA.  
 AC Q8BFL0  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Nitrite reductase (Fragment).  
 GN NIRK.  
 OS uncultured organism.  
 OC unclassified; environmental samples.  
 OX NCBI\_TaxID=155900;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=H02-05-118;  
 RA Fields M.W., Yan T., Wu L., Zu Y., Tiedje J.M., Zhou J.;  
 RT "Molecular diversity and characterization of nitrite reductase gene  
 RT fragments (nirK and nirS) from nitrate and uranium-contaminated  
 RT groundwater."  
 RL Environ. Microbiol. 0:0-0(2002).  
 DR EMBL: AF548945; AAN59868.1; -.  
 DR NON\_TER 1  
 FT NON\_TER 1  
 SQ SEQUENCE 102 AA; 11345 MW; 0909B399BC09E187 CRC64;

Query Match 2.2%; Score 11; DB 14; Length 102;  
 Best Local Similarity 100.0%; Pred. No. 0.0098;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 252 VGALTGENALK 262  
 Db 63 VGALTGENALK 73

## RESULT 23

Q8BPK6 PRELIMINARY; PRT; 102 AA.  
 AC Q8BPK6  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Nitrite reductase (Fragment).  
 GN NIRK.  
 OS uncultured organism.  
 OC unclassified; environmental samples.  
 OX NCBI\_TaxID=155900;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=G02-05-110;  
 RA Fields M.W., Yan T., Wu L., Zu Y., Tiedje J.M., Zhou J.;  
 RT "Molecular diversity and characterization of nitrite reductase gene  
 RT fragments (nirK and nirS) from nitrate and uranium-contaminated  
 RT groundwater."  
 RL Environ. Microbiol. 0:0-0(2002).  
 DR EMBL: AF548950; AAN59873.1; -.  
 DR NON\_TER 1  
 FT NON\_TER 1  
 SQ SEQUENCE 102 AA; 11264 MW; 45B54C96EEDD98D4 CRC64;

Query Match 2.2%; Score 11; DB 14; Length 102;  
 Best Local Similarity 100.0%; Pred. No. 0.0098;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 253 GALTGENALKA 263  
 Db 64 GALTGENALKA 74

## RESULT 24

ID Q9P9H9 PRELIMINARY; PRT: 361 AA.

AC Q9P9H9; 01-OCT-2000 (TReMBLrel. 15, Created)  
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)  
DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)  
DE Copper-containing dissimilatory nitrite reductase precursor  
(EC 1.7.99.3).

GN NTRK.  
OS Halorcula marismortui (Halobacterium marismortui).  
OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;  
OC Halobacteriaceae; Halorcula.  
OX NCBI\_TaxID=2238;  
RN [1]

RP SEQUENCE FROM N.A.  
RX MEDLINE=21311739; PubMed=11418554;  
RA Ichiki H., Tanaka Y., Mochizuki K., Yoshimatsu K., Sakurai T.,  
Fujiwara T.;  
RT "Purification, characterization, and genetic analysis of Cu-containing  
RT dissimilatory nitrite reductase from a denitrifying halophilic  
archaeon, Halorcula marismortui.";  
RL J. Bacteriol. 183:4149-4156(2001).  
DR EMBL; AJ278286; CAB93142.1; -.  
DR HSSP; P38501; IAS8.  
DR InterPro; IPR001117; Cu-oxidase.  
DR InterPro; IPR001287; CUNO2\_reductase.  
DR Pfam; PF00394; Cu-oxidase; 1.  
DR PRINTS; PR00695; CUNO2REDTASE.  
KW Oxidoreductase; Signal.  
FT CHAIN 1 31  
FT SIGNAL 1 31

FT CHAIN 32 361  
FT SIGNAL 32 361  
FT CHAIN 32 361  
FT SIGNAL 32 361

FT CHAIN 32 361  
FT SIGNAL 32 361  
FT CHAIN 32 361  
FT SIGNAL 32 361

FT CHAIN 32 361  
FT SIGNAL 32 361  
FT CHAIN 32 361  
FT SIGNAL 32 361

FT CHAIN 32 361  
FT SIGNAL 32 361  
FT CHAIN 32 361  
FT SIGNAL 32 361

FT CHAIN 32 361  
FT SIGNAL 32 361  
FT CHAIN 32 361  
FT SIGNAL 32 361

FT CHAIN 32 361  
FT SIGNAL 32 361  
FT CHAIN 32 361  
FT SIGNAL 32 361

FT CHAIN 32 361  
FT SIGNAL 32 361  
FT CHAIN 32 361  
FT SIGNAL 32 361

FT CHAIN 32 361  
FT SIGNAL 32 361  
FT CHAIN 32 361  
FT SIGNAL 32 361

FT CHAIN 32 361  
FT SIGNAL 32 361  
FT CHAIN 32 361  
FT SIGNAL 32 361

FT CHAIN 32 361  
FT SIGNAL 32 361  
FT CHAIN 32 361  
FT SIGNAL 32 361

FT CHAIN 32 361  
FT SIGNAL 32 361  
FT CHAIN 32 361  
FT SIGNAL 32 361

FT CHAIN 32 361  
FT SIGNAL 32 361  
FT CHAIN 32 361  
FT SIGNAL 32 361

FT CHAIN 32 361  
FT SIGNAL 32 361  
FT CHAIN 32 361  
FT SIGNAL 32 361

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 250 GSVGALTGE 258  
DB 83 GSVGALTGE 91

## RESULT 26

ID Q8XB7 PRELIMINARY; PRT: 428 AA.

AC Q8XB7; 01-MAR-2003 (TReMBLrel. 23, Created)  
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)  
DE Trigger factor (Prolyl isomerase).  
GN Q8078.  
OS Oceanobacillus iheyensis.  
OC Bacteria; Firmicutes; Bacillales; Oceanobacillus.  
OX NCBI\_TaxID=182710;  
RN [1]

RP SEQUENCE FROM N.A.  
RX STRAIN=HTE831 / DSM 14371 / JCM 11309;  
MEDLINE=22220767; PubMed=12335376;  
RA Takami H., Takaki Y., Uchiyama I.;  
RT "Genome sequence of Oceanobacillus iheyensis isolated from the Iheya  
RT Ridge and its unexpected adaptive capabilities to extreme  
RT environments.";  
RL Nucleic Acids Res. 30:3927-3935(2002).  
DR EMBL; AF004600; BAC14034.1; -.  
KW isomerase; Complete proteome.  
FT CHAIN 428 428  
FT SIGNAL 428 428

FT CHAIN 428 428  
FT SIGNAL 428 428  
FT CHAIN 428 428  
FT SIGNAL 428 428

FT CHAIN 428 428  
FT SIGNAL 428 428  
FT CHAIN 428 428  
FT SIGNAL 428 428

FT CHAIN 428 428  
FT SIGNAL 428 428  
FT CHAIN 428 428  
FT SIGNAL 428 428

FT CHAIN 428 428  
FT SIGNAL 428 428  
FT CHAIN 428 428  
FT SIGNAL 428 428

FT CHAIN 428 428  
FT SIGNAL 428 428  
FT CHAIN 428 428  
FT SIGNAL 428 428

FT CHAIN 428 428  
FT SIGNAL 428 428  
FT CHAIN 428 428  
FT SIGNAL 428 428

FT CHAIN 428 428  
FT SIGNAL 428 428  
FT CHAIN 428 428  
FT SIGNAL 428 428

FT CHAIN 428 428  
FT SIGNAL 428 428  
FT CHAIN 428 428  
FT SIGNAL 428 428

FT CHAIN 428 428  
FT SIGNAL 428 428  
FT CHAIN 428 428  
FT SIGNAL 428 428

FT CHAIN 428 428  
FT SIGNAL 428 428  
FT CHAIN 428 428  
FT SIGNAL 428 428

FT CHAIN 428 428  
FT SIGNAL 428 428  
FT CHAIN 428 428  
FT SIGNAL 428 428

FT CHAIN 428 428  
FT SIGNAL 428 428  
FT CHAIN 428 428  
FT SIGNAL 428 428

RA Iida T., Takami H., Honda T., Sakakawa C., Ogasawara N., Yasunaga T.,  
RA Kubara S., Shiba T., Hattori M., Shingawa H.,  
RT "Complete genome sequence of enterohemorrhagic *Escherichia coli*  
RT O157:H7 and genomic comparison with a laboratory strain K-12.";  
RL DNA Res. 8:11-22(2001).  
DR EMBL: AE005195; AAC54498.1; -;  
DR EMBL: AP002550; BAB33621.1; -;  
KM Complete proteome.  
SQ SEQUENCE 134 AA; 14177 MW; 824F8209E07BB41 CRC64;

Query Match 1.6%; Score 8; DB 16; Length 134;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 ALMLSGCS 24  
|||||  
DB 10 ALMLSGCS 17

## RESULT 28

Q8KPJ1 PRELIMINARY; PRT; 171 AA.  
ID Q8KPJ1  
AC Q8KPJ1  
DT 01-OCT-2002 (TREMBlrel. 22, Created)  
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)  
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
DE Nitrite reductase (Fragment).  
GN NITRK.  
OS uncultured bacterium.  
OC Bacteria; environmental samples.  
OK NCBI\_TaxID=77133;  
RN (1)  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21914100; PubMed=11916709;  
RA Prieme A., Braker G., Tiedje J.M.;  
RT "Diversity of nitrite reductase (nitrK and nirs) gene fragments in  
RT forested upland and wetland soils.";  
RL Appl. Environ. Microbiol. 68:1893-1900(2002).  
DR EMBL: AY121567; AAM94098.1; -;  
DR InterPro: IPR001117; Cu-oxidase.  
DR InterPro: IPR002160; Kunitz\_legume.  
DR Pfam: PF00394; Cu-oxidase; 1.  
DR PROSITE: PS00283; SOYBEAN\_KUNITZ; 1.  
FT NON\_TER 1 171  
SQ SEQUENCE 171 AA; 18917 MW; 47E538133E7D9317 CRC64;

Query Match 1.6%; Score 8; DB 2; Length 171;  
Best Local Similarity 100.0%; Pred. No. 18;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 246 VVFNCSVG 253  
|||||  
DB 79 VVFNCSVG 86

## RESULT 29

Q8KPM6 PRELIMINARY; PRT; 172 AA.  
ID Q8KPM6  
AC Q8KPM6  
DT 01-OCT-2002 (TREMBlrel. 22, Created)  
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)  
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
DE Nitrite reductase (Fragment).  
GN NITRK.  
OS uncultured bacterium.  
OC Bacteria; environmental samples.  
OK NCBI\_TaxID=77133;  
RN (1)  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21914100; PubMed=11916709;  
RA Prieme A., Braker G., Tiedje J.M.;  
RT "Diversity of nitrite reductase (nitrK and nirs) gene fragments in

RT forested upland and wetland soils.";

RT forested upland and wetland soils.";  
RL Appl. Environ. Microbiol. 68:1893-1900(2002).  
DR EMBL: AY121527; AAM94058.1; -;  
DR InterPro: IPR001117; Cu-oxidase.  
DR Pfam: PF00394; Cu-oxidase; 1.  
FT NON\_TER 1 172  
FT NON\_TER 1 172  
SQ SEQUENCE 172 AA; 19069 MW; B477803BA4FCF50F CRC64;

Query Match 1.6%; Score 8; DB 2; Length 172;  
Best Local Similarity 100.0%; Pred. No. 19;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 255 LTGENALK 262  
|||||  
DB 88 LTGENALK 95

## RESULT 30

Q8TSG4 PRELIMINARY; PRT; 252 AA.  
ID Q8TSG4  
AC Q8TSG4  
DT 01-JUN-2002 (TREMBlrel. 21, Created)  
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)  
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
DE Formylmethanofuran dehydrogenase, subunit C.  
GN FMDC OR MA0832.  
OS Methanosarcina acetivorans.  
OC Archaea; Euryarchaeota; Methanococci; Methanosarcinales;  
OC Methanosarcinaceae; Methanosarcina.  
OK NCBI\_TaxID=2214;  
RN (1)  
RP SEQUENCE FROM N.A.  
RX STRAIN=C2A / ATCC 35395 / DSM 2834;  
RX MEDLINE=21929760; PubMed=11932238;  
RA Galagan J.E., Nussbaum C., Roy A., Endrizzi M.G., Macdonald P.,  
RA Fitzhugh W., Calvo S., Engels R., Smirnov S., Alnoor D., Brown A.,  
RA Allen N., Naylor J., Stange-Thomann N., Deatellano K., Johnson R.,  
RA Linson L., McEwan P., McKernan K., Talamas J., Tirrell A., Ye W.,  
RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame D.A., Guse A.M.,  
RA Hederich R., Ingram-Smith C., Kuettner H.C., Krzycki J.A.,  
RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,  
RA Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,  
RA Ferry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.,  
RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,  
RA Metcalf W.W., Birren B.;  
RT "The genome of *Methanosarcina acetivorans* reveals extensive metabolic  
RT and physiological diversity.";  
RL Genome Res. 12:532-542(2002).  
DR EMBL: AE010745; AAM04271.1; -;  
DR InterPro: IPR002489; DUF14.  
DR Pfam: PF01493; GKGXG; 1.  
KM Complete proteome.  
SQ SEQUENCE 252 AA; 26676 MW; 91929A25282421B1 CRC64;

Query Match 1.6%; Score 8; DB 17; Length 252;  
Best Local Similarity 100.0%; Pred. No. 27;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 447 GKITVNGN 454  
|||||  
DB 150 GKITVNGN 157

## RESULT 31

Q8PVI5 PRELIMINARY; PRT; 252 AA.  
ID Q8PVI5  
AC Q8PVI5  
DT 01-OCT-2002 (TREMBlrel. 22, Created)  
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)  
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
DE Tungsten formylmethanofuran dehydrogenase, subunit C (EC  
1.2.99.5);

GN MM1978.  
 OS Methanosarcina mazel (Methanosarcina frisia).  
 OC Archaea; Euryarchaeota; Methanococci; Methanosarcinales;  
 OC Methanosarcinaceae; Methanosarcina.  
 OX NCBI\_TaxID=2209;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Goel / Go1 / ATCC BAA-199 / DSM 3647 / OCM 88;  
 RX MEDLINE=2120827; PubMed=12125824;  
 RA Deppenmeier U., Johann A., Hartesch T., Merkl R., Schmitz R.A.,  
 RA Martinez-Arias R., Henne A., Wieser A., Baumer S., Jacobi C.,  
 RA Brueggemann H., Lienack T., Christmann A., Boemcke M., Steckel S.,  
 RA Bhattacharyya A., Lykidis A., Overbeek R., Klenk H.-P., Gunsalus R.P.,  
 RA Fritz H.-J., Gottschalk G.;  
 RT "The genome of Methanosarcina mazel: evidence for lateral gene  
 RT transfer between Bacteria and Archaea.";  
 RL J. Mol. Microbiol. Biotechnol. 4:453-461(2002).  
 DR EMBL; AB013435; AAM31674.1; -;  
 DR InterPro; IPR002489; DUF14.  
 DR Pfam; PF01493; GXXGX; 1.  
 KW Oxidoreductase; Complete proteome.  
 SO SEQUENCE 252 AA; 26742 MW; 59CA399B1BBELAC CRC64;

Query Match 1.6%; Score 8; DB 17; Length 252;  
 Best Local Similarity 100.0%; Pred. No. 27;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 447 GKITVNGN 454  
 DB 150 GKITVNGN 157

## RESULT 32

Q9FFM7. PRELIMINARY; PRT; 255 AA.

AC Q9FFM7;  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE ABC transporter-like protein (Fragment).  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosid II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Columbia;  
 RX MEDLINE=97471969; PubMed=9330910;  
 RA Sato S., Kotani H., Nakamura Y., Kaneko T., Asamizu E., Fukami M.,  
 RA Miyajima N., Tabata S.;  
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. I. Sequence  
 RT features of the 1.6 Mb regions covered by twenty physically assigned  
 RT clones.";  
 RL DNA Res. 4:215-230(1997).  
 CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.  
 DR EMBL; AB005235; BAB17024.1; -;  
 DR InterPro; IPR003593; AAA\_ATPase.  
 DR InterPro; IPR003439; ABC\_transporter.  
 DR Pfam; PF00005; ABC\_tran; 1.  
 DR ProDom; PD000006; ABC\_transporter; 1.  
 DR SMART; SM00382; AAA; 1.  
 DR PROSITE; PS00211; ABC\_TRANSPORTER; 1.  
 KW ATP-binding; Transport.  
 FT NON\_TER 1  
 SQ SEQUENCE 255 AA; 27832 MW; 79BDA0BD5378BCD5 CRC64;

Query Match 1.6%; Score 8; DB 10; Length 255;  
 Best Local Similarity 100.0%; Pred. No. 27;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 AAKTANAD 48  
 DB 150 AAKTANAD 157

DB 129 AAKTANAD 136

## RESULT 33

Q8EU09. PRELIMINARY; PRT; 321 AA.

AC Q8EU09;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE O-sialoglycoprotein end peptidase.  
 OS Myxobolus penicillatus.  
 OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.  
 OX NCBI\_TaxID=28227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=HF-2;  
 RX MEDLINE=22354719; PubMed=12465555;  
 RA Sasaki Y., Ishikawa J., Yamashita A., Oshima K., Kenri T., Furuya K.,  
 RA Yoshino C., Horino A., Shiba T., Sasaki T., Hattori M.;  
 RT "The complete genomic sequence of Mycoplasma penicillatus, an  
 RT intracellular bacterial pathogen in humans.";  
 RL Nucleic Acids Res. 30:5293-5300(2002).  
 DR EMBL; AP004173; BAC44653.1; -;  
 KW Complete proteome.  
 SO SEQUENCE 321 AA; 35892 MW; EFP22FCA4DBB0BDE CRC64;

Query Match 1.6%; Score 8; DB 16; Length 321;  
 Best Local Similarity 100.0%; Pred. No. 33;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 287 IGEIFDKV 294  
 DB 175 IGEIFDKV 182

## RESULT 34

O68601. PRELIMINARY; PRT; 360 AA.

AC O68601;  
 DT 01-AUG-1998 (TrEMBLrel. 07, Created)  
 DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
 DE Dissimilatory copper-containing nitrite reductase.  
 GN NTR.  
 OS Alcaligenes xyloxydans xyloxydans (Achromobacter xyloxydans).  
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
 OC Alcaligenaceae; Achromobacter.  
 OX NCBI\_TaxID=515;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NCIMB11015;  
 RA Suzuki E., Horikoshi N., Kohzuma T.;  
 RL Submitted (MAR-1998) to the EMBL/Genbank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GF1051;  
 RA Kataoka K., Furusawa H., Yamaguchi K., Suzuki S.;  
 RT "Cloning and Expression of Copper Nitrite Reductase Gene from  
 RT Alcaligenes xyloxydans GF1051.";  
 RL Submitted (APR-1998) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; AF051831; AAC05831.1; -;  
 DR EMBL; AB013078; BAA33678.1; -;  
 DR HSSP; P25006; INIF.  
 DR InterPro; IPR001117; Cu-oxidase.  
 DR InterPro; IPR001287; CuNO2\_reductase.  
 DR Pfam; PF00394; Cu-oxidase; 2.  
 DR PRINTS; PR00695; CUNO2REDTASE.  
 KW Signal.  
 SQ SEQUENCE 360 AA; 38939 MW; 3748B5BD3FFP44E7 CRC64;

Query Match 1.6%; Score 8; DB 2; Length 360;

Best Local Similarity 100.0%; Pred. No. 37;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 137 MFHNVDFH 144  
|||||  
Db 111 MFHNVDFH 118

## RESULT 35

Q9XK1C4 PRELIMINARY; PRT; 483 AA.  
AC Q9XK1C4;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
GN Hypothetical protein NMB0240.  
OS Neisseria meningitidis (serogroup B).  
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;  
OC Neisseriaceae; Neisseria.  
OX NCBI\_TaxID=491;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MC58 / Serogroup B;  
RX MEDLINE=20175755; PubMed=10710307;  
RA Tetselin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,  
RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,  
RA Nelson W.C., Gwin M.L., Deboy R., Peterson J.D., Hickey E.K.,  
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,  
RA Mason T., Clecko A., Parksey D.S., Blair E., Clifton H., Clark E.B.,  
RA Cotton M.D., Uterback T.R., Khouri H., Qin H., Vamathevan J.,  
RA Gill J., Scarlato V., Maignani V., Pizsa M., Grandi G., Sun L.,  
RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;  
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain  
MC58.";  
RT Science 287:1809-1815(2000).  
DR EMBL; AE002381; AAF40694.1; -;  
DR TIGR; NMB0240; -;  
DR InterPro; IPR000051; SAM bind.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 483 AA; 53511 MW; 53C914D86F5D097 CRC64;

Query Match 1.6%; Score 8; DB 16; Length 483;  
Best Local Similarity 100.0%; Pred. No. 49;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 35 SSTVDAAA 42  
|||||  
Db 440 SSTVDAAA 447

## RESULT 36

Q9JX77 PRELIMINARY; PRT; 494 AA.  
AC Q9JX77;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE Putative integral membrane protein.  
GN NMA0020.  
OS Neisseria meningitidis (serogroup A).  
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;  
OC Neisseriaceae; Neisseria.  
OX NCBI\_TaxID=56599;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=22491 / Serogroup A / Serotype 4A;  
RX MEDLINE=2022556; PubMed=10761919;  
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,  
RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,  
RA Davis R.M., Davis P., Devlin K., Fellwell T., Hamlin N., Holtroyd S.,  
RA Jagels K., Leather S., Moule S., Mungall K., Quail M.A.,  
RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,

RA Whitehead S., Spratt B.G., Barrell B.G.;  
RT "Complete DNA sequence of a serogroup A strain of Neisseria  
RT meningitidis 22491."  
RL Nature 404:502-506(2000).  
DR EMBL; AL162752; CAB83340.1; -;  
DR InterPro; IPR000051; SAM\_bind.  
KW Complete proteome.  
SQ SEQUENCE 494 AA; 54774 MW; FB45E96837631186 CRC64;

Query Match 1.6%; Score 8; DB 16; Length 494;  
Best Local Similarity 100.0%; Pred. No. 50;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 35 SSTVDAAA 42  
|||||  
Db 451 SSTVDAAA 458

## RESULT 37

Q98112 PRELIMINARY; PRT; 509 AA.  
AC Q98112;  
DT 01-OCT-2001 (TrEMBLrel. 18, Created)  
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)  
DE 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
GN Probable oxidoreductase.  
ML M2617.  
OS Rhizobium loti (Mesorhizobium loti).  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
OC Phyllobacteriaceae; Mesorhizobium.  
OX NCBI\_TaxID=381;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MAFP303099;  
RX MEDLINE=21082930; PubMed=11214968;  
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,  
RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,  
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,  
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,  
RA Takeuchi C., Yamada M., Tabata S.;  
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium  
Mesorhizobium loti."  
RT DNA Res. 7:331-338(2000).  
RL EMBL; AP003000; BAB49704.1; -;  
DR InterPro; IPR006094; Oxid\_FAD\_bind.  
DR Pfam; PF01565; FAD\_binding\_4; 1.  
KW Complete proteome.  
SQ SEQUENCE 509 AA; 55729 MW; E57EDD2155129CAF CRC64;

Query Match 1.6%; Score 8; DB 16; Length 509;  
Best Local Similarity 100.0%; Pred. No. 51;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 37 TVDAAKT 44  
|||||  
Db 116 TVDAAKT 123

## RESULT 38

Q9AV50 PRELIMINARY; PRT; 514 AA.  
AC Q9AV50;  
DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
DE Putative extensin.  
GN OSUNBA0093B11.1.  
OS Oryza sativa (Rice).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
OC Eriarthroidae; Mynoliophyta; Liliopsida; Poales; Poaceae;  
OX NCBI\_TaxID=4530;  
RN [1]

RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Nipponbare;  
 RA Bell C.R., Yuan O., Ouyang S., Moffat K.S., Hill J.N., Burr P.C.,  
 RA Heiao J., Ziesman V., Pat G., Bowman C.L., Fujii C.Y., Vanken S.E.,  
 RA Bowman C.L., Craven B., Uteback T.R., Kalak H., Feldlyum I.V.,  
 RA Quackenbush J., White O., Salzberg S.L., Fraser C.M.;  
 RT "Olyta sativa chromosome 10 BAC OSUBa0093b1 genomic sequence."  
 RT Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AC024594; AAK21352.1; -  
 DR HSSP: Q91835; 1D12.  
 DR Gramene: Q9AV50; -  
 DR InterPro: IPR001159; DS\_RBD.  
 DR InterPro: IPR002965; P\_Rich\_extensn.  
 DR Pfam: PF00035; dsm; 2.  
 DR PRINTS: PRO1217; PRICHEXTNSN.  
 DR SMART: SM00358; DSRM; 2.  
 DR PROSITE: PS50137; DS\_RBD; 2.  
 SQ SEQUENCE 514 AA; 56078 MW; A98AD36128502939 CRC64;

Query Match 1.6%; Score 8; DB 10; Length 514;  
 Best Local Similarity 100.0%; Pred. No. 52;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 43 KTANADNA 50  
 DB 483 KTANADNA 490

RESULT 39  
 ID Q91ZB8 PRELIMINARY; PRT; 634 AA.  
 AC Q91ZB8;  
 DT 01-OCT-2000 (TEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TEMBLrel. 15, Last sequence update)  
 DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)  
 DE ABC transporter-like protein (Putative ABC transporter protein).  
 GN F8F6\_120 OR AT5G03910.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC euroids II; Brassicales; Brassicaceae; Arabidopsis.  
 OC NCBI\_TaxID=3702;  
 RX [1]  
 RP SEQUENCE FROM N.A.  
 RA Benav M., Peters S.A., van Staveren M., Dirkee W., Stiekema W.,  
 RA Bancroft I., Mewes H.W., Rudd S., Lemcke K., Mayer K.F.X.;  
 RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA EU Arabidopsis sequencing project;  
 RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Yamada K., Banh J., Banno F., Chang E., Dale J.M., Goldsmith A.D.,  
 RA Lee J.M., Onodera C.S., Quach H.L., Tang C., Toriumi M., Wu H.C.,  
 RA Yamamura Y., Yu G., Yu S., Bowser L., Carninci P., Chen H., Cheuk R.,  
 RA Hayashizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,  
 RA Kawai J., Kim C., Koeseema E., Lam B., Lin J., Meyers M.C., Miranda M.,  
 RA Narusaka M., Nguyen W., Palm C.U., Sakurai T., Saitou M., Seki M.,  
 RA Shinn P., Southwick A., Tracy S.E., Shinzaki K., Davis R.W.,  
 RA Ecker J.R., Theologis A.;  
 RT "Full length cDNA of gene At5G03910 (GI:15237574).";  
 RT Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Yamada K., Banh J., Chan M.M., Chang C.H., Chang E., Dale J.M.,  
 RA Deng J.M., Goldsmith A.D., Lee J.M., Onodera C.S., Quach H.L.,  
 RA Tang C., Toriumi M., Wu H.C., Yamamura Y., Yu G., Bowser L.,  
 RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,  
 RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,  
 RA Meyers M.C., Miranda M., Narusaka M., Nguyen M., Nguyen M., Palm C.U.,  
 RA Sakurai T., Saitou M., Seki M., Shinn P., Southwick A., Shinzaki K.,  
 RA Davis R.W., Ecker J.R., Theologis A.;

RT "Arabidopsis Open Reading Frame (ORF) Clones."  
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.  
 DR EMBL: AL162873; CAB8511.1; -  
 DR EMBL: AY059727; AAL24084.1; -  
 DR EMBL: AY059727; AAL24084.1; -  
 DR InterPro: IPR003593; AAA\_ATPase.  
 DR InterPro: IPR001140; ABC\_TM\_transp.  
 DR InterPro: IPR003439; ABC\_transporter.  
 DR Pfam: PF00664; ABC\_membrane; 1.  
 DR Pfam: PF00005; ABC\_tran; 1.  
 DR ProDom: PD000006; ABC\_transporter; 1.  
 DR SMART: SM00382; AAA; 1.  
 DR PROSITE: PS00211; ABC\_TRANSPORTER; 1.  
 DR ATP-binding; Transport.  
 SQ SEQUENCE 634 AA; 69197 MW; 957B5456221871A3 CRC64;

Query Match 1.6%; Score 8; DB 10; Length 634;  
 Best Local Similarity 100.0%; Pred. No. 63;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 AAKTANAD 48  
 DB 508 AAKTANAD 515

RESULT 40  
 ID Q93538 PRELIMINARY; PRT; 915 AA.  
 AC Q93538;  
 DT 01-FEB-1997 (TEMBLrel. 02, Created)  
 DT 01-FEB-1997 (TEMBLrel. 02, Last sequence update)  
 DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)  
 DE F20D1.6 protein.  
 GN F20D1.6.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;  
 OC Rhabditidae; Pelodierinae; Caenorhabditis.  
 OC NCBI\_TaxID=6239;  
 RX [1]  
 RP SEQUENCE FROM N.A.  
 RA Burton J.;  
 RA Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=99069613; PubMed=9851916;  
 RX none;  
 RA "Genome sequence of the nematode C.elegans: A platform for  
 RT investigating biology."  
 RT Science 282:2012-2018(1998).  
 DR EMBL: Z78542; CAB01748.1; -  
 DR WormRep; F20D1.6; CE09499.  
 SQ SEQUENCE 915 AA; 104585 MW; 0AEAF6A3A58F8566 CRC64;

Query Match 1.6%; Score 8; DB 5; Length 915;  
 Best Local Similarity 100.0%; Pred. No. 88;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 87 KMETVEKV 94  
 DB 503 KMETVEKV 510

Search completed: August 27, 2003, 18:45:12  
 Job time : 143 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 27, 2003, 18:34:22 ; Search time 42 Seconds  
(without alignments)  
1149.445 Million cell updates/sec

Title: US-10-088-045-2

Perfect score: 502  
Sequence: 1 MSKPTLIKTTLCALSLML.....NKGQLSADDVAKAKTKRN 502

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 283308 seqs, 96168682 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database : PIR\_76: \*  
1: pir1: \*  
2: pir2: \*  
3: pir3: \*  
4: pir4: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	34	6.8	386	2 B81816	nitrite reductase
2	34	6.8	390	2 B81062	nitrite reductase
3	34	6.8	392	2 A49208	nitrite reductase
4	12	2.4	363	2 S32112	(EC 1.7.2.1) (EC
5	8	1.6	134	2 D64744	exopolysaccharide
6	8	1.6	134	2 F90653	regulator in colan
7	8	1.6	134	2 F85504	regulator in colan
8	8	1.6	336	2 JE0215	nitrite reductase
9	8	1.6	360	2 JG0170	nitrite reductase
10	8	1.6	389	2 D64333	pyruvate synthase
11	8	1.6	483	2 F81221	hypothetical prote
12	8	1.6	494	2 H81592	probable integral
13	8	1.6	604	2 F69802	ABC transporter (A
14	8	1.6	634	2 T48418	ABC transporter (A
15	8	1.6	915	2 T21147	hypothetical prote
16	8	1.4	86	1 CCCP55	cytochrome c555 -
17	7	1.4	98	2 AG3064	hypothetical prote
18	7	1.4	104	2 S52930	GP41 ENV protein -
19	7	1.4	109	2 S16437	ribulose-bisphosph
20	7	1.4	111	2 A70407	hypothetical prote
21	7	1.4	110	2 E86906	ribosomal protein
22	7	1.4	157	2 AF3315	transcription elon
23	7	1.4	161	2 B97802	hypothetical prote
24	7	1.4	175	2 A97742	cytochrome c limpo
25	7	1.4	191	2 AH0196	probable lipoprote
26	7	1.4	192	2 T29645	hypothetical prote
27	7	1.4	193	1 F64941	rnd protein - Esch
28	7	1.4	193	2 C90943	probable outer mem
29	7	1.4	193	2 G85791	probable outer mem

30	7	1.4	193	2 AD0725	probable lipoprote
31	7	1.4	194	2 G69539	ribosomal protein
32	7	1.4	202	2 T36138	hypothetical prote
33	7	1.4	211	2 A85098	hypothetical prote
34	7	1.4	211	2 C84775	probable harpin-in
35	7	1.4	212	2 C70737	probable sigd prot
36	7	1.4	212	2 T29906	hypothetical prote
37	7	1.4	215	2 S55925	hypothetical arabinoga
38	7	1.4	218	2 G91207	probable replicase
39	7	1.4	230	2 T15381	hypothetical prote
40	7	1.4	238	2 I38849	LEK-3 - human
41	7	1.4	239	2 S64327	probable membrane
42	7	1.4	245	2 B86889	zinc ABC transport
43	7	1.4	255	2 G82644	2,5-dichloro-2,5-c
44	7	1.4	258	2 AC2860	d-beta-hydroxybuty
45	7	1.4	261	2 A83196	probable transcrip
46	7	1.4	282	2 AG3047	transcription regu
47	7	1.4	282	2 AB2037	hypothetical prote
48	7	1.4	293	2 D86670	pseudouridine synt
49	7	1.4	293	2 T29907	hypothetical prote
50	7	1.4	296	2 T46617	probable chemotaxi
51	7	1.4	299	2 T52444	hypothetical prote
52	7	1.4	299	2 AC2998	glutamy-tRNA synt
53	7	1.4	299	2 D98298	glutamy-tRNA synt
54	7	1.4	299	2 F98285	helix-turn-helix d
55	7	1.4	305	2 A97637	glutamy-tRNA synt
56	7	1.4	309	2 A81403	D-beta-hydroxybuty
57	7	1.4	312	2 F86044	probable periplasm
58	7	1.4	312	2 B91198	secreted protein E
59	7	1.4	313	2 D96616	Esap protein limpo
60	7	1.4	315	2 C82933	hypothetical prote
61	7	1.4	320	2 AD3435	pseudouridine synt
62	7	1.4	322	2 G97317	cytochrome-c oxida
63	7	1.4	323	2 D87472	probable dehydroge
64	7	1.4	325	2 A86054	hypothetical prote
65	7	1.4	326	2 S66262	probable replicase
66	7	1.4	327	2 T07104	vestitone reductas
67	7	1.4	331	2 S60671	2'-hydroxydihydrod
68	7	1.4	334	2 B69481	ParB-like partitio
69	7	1.4	338	2 D97166	hypothetical prote
70	7	1.4	341	2 A87293	flagellar motor sw
71	7	1.4	341	2 A83644	hypothetical prote
72	7	1.4	352	2 B43649	hypothetical prote
73	7	1.4	359	2 T22774	hypothetical prote
74	7	1.4	365	2 B48945	recombination prot
75	7	1.4	373	2 G75073	hypothetical prote
76	7	1.4	373	2 A86798	prophage p13 prote
77	7	1.4	376	2 I39582	nitrite reductase
78	7	1.4	378	2 AC3633	nitrite reductase
79	7	1.4	378	2 JC4648	nitrite reductase
80	7	1.4	378	2 E70786	hypothetical prote
81	7	1.4	379	2 T31154	hypothetical prote
82	7	1.4	387	2 B86669	Reca protein limpo
83	7	1.4	391	2 C72220	conserved hypochet
84	7	1.4	393	2 S38875	methionine adenosy
85	7	1.4	397	2 S66352	methionine adenosy
86	7	1.4	411	2 AB3095	nitrite reductase,
87	7	1.4	420	2 H75395	ABC transporter,
88	7	1.4	436	2 B55452	cardiocyte-derived
89	7	1.4	438	2 E83938	allantoinase BH230
90	7	1.4	438	2 C89864	coenzyme A disulfid
91	7	1.4	444	2 T03566	probable trigiger f
92	7	1.4	450	2 D36953	dihydroliipoamide S
93	7	1.4	455	2 H98191	copper-containing
94	7	1.4	459	2 S64741	hypothetical prote
95	7	1.4	462	2 AF1694	3-isopropylmalate
96	7	1.4	472	1 A49340	membrane-bound alc
97	7	1.4	495	2 A95984	probable xanthine
98	7	1.4	501	2 D87613	pilus assembly pro
99	7	1.4	502	2 D84119	ATP synthase alpha
100	7	1.4	512	2 S19354	cell fusion protei
101	7	1.4	516	2 A95134	voltage-gated chlo
102	7	1.4	521	2 A29345	steroid hormone re

probable lipoprote  
ribosomal protein  
hypothetical prote  
hypothetical prote  
probable harpin-in  
probable sigd prot  
hypothetical prote  
hypothetical arabinoga  
probable replicase  
hypothetical prote  
LEK-3 - human  
probable membrane  
zinc ABC transport  
2,5-dichloro-2,5-c  
d-beta-hydroxybuty  
probable transcrip  
transcription regu  
hypothetical prote  
pseudouridine synt  
hypothetical prote  
probable chemotaxi  
hypothetical prote  
glutamy-tRNA synt  
glutamy-tRNA synt  
helix-turn-helix d  
glutamy-tRNA synt  
D-beta-hydroxybuty  
probable periplasm  
secreted protein E  
Esap protein limpo  
hypothetical prote  
pseudouridine synt  
cytochrome-c oxida  
probable dehydroge  
hypothetical prote  
probable replicase  
vestitone reductas  
2'-hydroxydihydrod  
ParB-like partitio  
hypothetical prote  
flagellar motor sw  
hypothetical prote  
hypothetical prote  
hypothetical prote  
hypothetical prote  
recombination prot  
hypothetical prote  
prophage p13 prote  
nitrite reductase  
nitrite reductase  
nitrite reductase  
hypothetical prote  
hypothetical prote  
Reca protein limpo  
conserved hypochet  
methionine adenosy  
methionine adenosy  
nitrite reductase,  
ABC transporter,  
cardiocyte-derived  
allantoinase BH230  
coenzyme A disulfid  
probable trigiger f  
dihydroliipoamide S  
copper-containing  
hypothetical prote  
3-isopropylmalate  
membrane-bound alc  
probable xanthine  
pilus assembly pro  
ATP synthase alpha  
cell fusion protei  
voltage-gated chlo  
steroid hormone re

103	7	1.4	521	2	E98002	conserved hypochet	176	6	1.2	83	1	CCML6	cytochrome c6 - go
104	7	1.4	525	2	T06104	2-dehydro-3-deoxy-	177	6	1.2	83	2	FB9920	conserved hypochet
105	7	1.4	525	2	A41370	hypothetical prote	178	6	1.2	85	1	CCPR6	cytochrome c6 - re
106	7	1.4	543	2	T22585	diphosphate-fructo	179	6	1.2	85	1	CCPR6	cytochrome c6 - pl
107	7	1.4	544	2	S52081	probable amino aci	180	6	1.2	85	2	UCS849	cytochrome c6 - re
108	7	1.4	550	2	T37519	p-aminobenzoate sy	181	6	1.2	86	1	CCBR6	cytochrome c6 - ye
109	7	1.4	567	2	A71951	aspartyl-tRNA synt	182	6	1.2	87	1	CCYCG6	cytochrome c6 - sy
110	7	1.4	595	2	AH2718	aspartyl-tRNA synt	183	6	1.2	87	1	CCYCG6	cytochrome c6 - sy
111	7	1.4	595	2	D97500	probable very-long	184	6	1.2	88	1	CCBM6	cytochrome c6 - gr
112	7	1.4	608	2	H83284	glutamine-fructose	185	6	1.2	88	2	UC7029	cytochrome c6 - gr
113	7	1.4	611	2	F82951	transcription init	186	6	1.2	88	2	UCS816	cytochrome c6 - lval
114	7	1.4	615	2	S58230	hypothetical prote	187	6	1.2	88	2	A83659	hypothetical prote
115	7	1.4	633	2	S48956	conserved hypochet	188	6	1.2	89	1	CCSG6	cytochrome c6 - sp
116	7	1.4	663	2	F82160	phenol 2-monooxyge	189	6	1.2	89	1	S35677	cytochrome c6 - lval
117	7	1.4	664	2	A47012	hypothetical prote	190	6	1.2	89	2	S57923	cytochrome c6 - gr
118	7	1.4	677	2	T22534	methyalmalonyl-CoA	191	6	1.2	90	2	E96524	protein T1N15.5 (l
119	7	1.4	714	2	E87543	hypothetical prote	192	6	1.2	91	2	D86158	F22D16.18 protein
120	7	1.4	716	2	A75595	hypothetical prote	193	6	1.2	92	2	A69153	hypothetical prote
121	7	1.4	733	2	B40595	methyalmalonyl-CoA	194	6	1.2	93	2	G82603	conserved hypochet
122	7	1.4	744	2	I51022	replication licens	195	6	1.2	94	2	S30186	NADH2 dehydrogenas
123	7	1.4	747	2	A86659	anaerobic ribonuc	196	6	1.2	95	2	T06805	ribosomal protei
124	7	1.4	779	2	AE2402	alpha-glucosidase	197	6	1.2	95	2	T12723	hypothetical prote
125	7	1.4	807	1	I51685	replication licens	198	6	1.2	95	2	S77799	phosphoglycerate k
126	7	1.4	844	2	B64678	NADH2 dehydrogenas	199	6	1.2	97	2	AB1666	glutamyl-tRNA (Gln)
127	7	1.4	849	2	H71838	aluminum resistan	200	6	1.2	97	2	AD1294	glutamyl-tRNA (Gln)
128	7	1.4	859	2	S66827	serine/threonine p	201	6	1.2	98	2	S70250	ferredoxin 2(Fe-4
129	7	1.4	923	2	T08033	probable disease r	202	6	1.2	100	2	F72751	hypothetical prote
130	7	1.4	957	2	E84547	potassium channel	203	6	1.2	101	2	E72691	hypothetical prote
131	7	1.4	962	2	I53197	potassium channel	204	6	1.2	101	2	C90564	lipoprotein (lipor
132	7	1.4	989	2	I48912	protein C05D11.1 (	205	6	1.2	102	2	S06101	very hypochet
133	7	1.4	995	2	A88483	probable ATP-bindi	206	6	1.2	102	2	T40829	ribosomal protei
134	7	1.4	1030	2	S43211	cytoplasmic linker	207	6	1.2	104	2	H64225	nitrite reductase
135	7	1.4	1046	2	T42720	cytoplasmic linker	208	6	1.2	108	2	A65131	nitrite reductase
136	7	1.4	1046	2	T42720	creatine kinase (E	209	6	1.2	108	2	A91156	nitrite reductase
137	7	1.4	1174	1	A43736	potassium channel	210	6	1.2	108	2	F86001	nitrite reductase
138	7	1.4	1174	1	A40853	hypothetical prote	211	6	1.2	108	2	AE0020	nitrite reductase
139	7	1.4	1206	2	A64207	streptococcal hema	212	6	1.2	108	2	AF1001	nitrite reductase
140	7	1.4	1214	2	G97419	conserved hypochet	213	6	1.2	109	1	A55124	chloredoxin - Chlo
141	7	1.4	1248	2	AH2637	probable polyketid	214	6	1.2	109	2	T02716	acidic ribosomal p
142	7	1.4	1461	2	B70588	neurogenic locus m	215	6	1.2	109	2	T02039	acidic ribosomal p
143	7	1.4	1596	2	A33106	hypothetical prote	216	6	1.2	109	2	G71249	hypothetical prote
144	7	1.4	1940	2	F75393	vacuolar proteina s	217	6	1.2	109	2	T02762	anther specific pr
145	7	1.4	3071	2	T50345	riofamycin polyketi	218	6	1.2	110	1	S73121	cytochrome c6 prec
146	7	1.4	3413	2	T17467	metalloproteinase	219	6	1.2	111	1	J01083	cytochrome c553 -
147	6	1.2	20	2	A58903	antifreeze protein	220	6	1.2	112	1	S35440	acidic ribosomal p
148	6	1.2	33	1	FDF15G	antifreeze protein	221	6	1.2	112	2	S35440	acidic ribosomal p
149	6	1.2	33	2	A05162	urease accessory p	222	6	1.2	112	2	AH0303	probable membrane
150	6	1.2	42	2	A42558	hypothetical prote	223	6	1.2	113	2	A47756	acidic ribosomal p
151	6	1.2	42	2	G95868	a 2-2 protein - in	224	6	1.2	114	1	R5CH2E	hypochet
152	6	1.2	44	2	S52541	ATP synthase Fo ch	225	6	1.2	114	2	D75420	hypochet
153	6	1.2	50	2	A82759	hypothetical prote	226	6	1.2	114	2	B45036	pur beta - human (
154	6	1.2	55	2	E90630	light-harvesting p	227	6	1.2	115	2	F84931	ribonuclease p (EC
155	6	1.2	55	2	A82823	hypothetical prote	228	6	1.2	115	2	D84976	50S ribosomal prot
156	6	1.2	56	2	S61509	adipokinetic hormo	229	6	1.2	116	2	E83010	conserved hypochet
157	6	1.2	58	1	C72388	hypothetical prote	230	6	1.2	116	2	H71034	probable translati
158	6	1.2	61	1	AKLQ82	hypochet	231	6	1.2	116	2	E72639	hypochet
159	6	1.2	61	2	T13448	hypochet	232	6	1.2	117	2	C64094	ribosomal protei
160	6	1.2	67	2	A82175	hypochet	233	6	1.2	117	2	UC5753	ribosomal protei
161	6	1.2	70	2	A11950	hypochet	234	6	1.2	117	2	F82057	ribosomal protei
162	6	1.2	71	2	C84341	hypochet	235	6	1.2	118	2	T27846	cytochrome c6 prec
163	6	1.2	73	2	AC0739	probable cell divi	236	6	1.2	120	1	A53328	hypochet
164	6	1.2	77	2	H69420	hydrogenase expres	237	6	1.2	120	2	T02708	hypochet
165	6	1.2	77	2	H90870	hypochet	238	6	1.2	120	2	T20150	hypochet
166	6	1.2	77	2	A85748	unknown protein en	239	6	1.2	120	2	B71875	hypochet
167	6	1.2	77	2	C64885	cell division inh1	240	6	1.2	120	2	B71875	hypochet
168	6	1.2	78	2	S44993	gag polyprotein -	241	6	1.2	123	2	A83041	conserved hypochet
169	6	1.2	78	2	AG1071	conserved hypochet	242	6	1.2	124	2	A75374	hypochet
170	6	1.2	79	2	T10336	hypochet	243	6	1.2	125	2	E97414	hypochet
171	6	1.2	80	2	H91230	hypochet	244	6	1.2	126	2	B82715	50S ribosomal prot
172	6	1.2	80	2	G86077	hypochet	245	6	1.2	126	2	A83890	hypochet
173	6	1.2	80	2	S40834	hypochet	246	6	1.2	126	2	AE2069	hypochet
174	6	1.2	80	2	AH3241	hypochet	247	6	1.2	127	2	F97157	stage III sporulat
175	6	1.2	82	2	AG0946	probable CopG-fam1	248	6	1.2	127	2	A84846	hypochet

249	6	1.2	129	2	C83113	50S ribosomal prot	322	6	1.2	162	2	F86637	hypothetical prote
250	6	1.2	129	2	T10175	acyl carrier prote	323	6	1.2	163	2	F72565	hypothetical prote
251	6	1.2	129	2	T07643	PEARL1 protein h	324	6	1.2	164	2	A26490	histone H1, macron
252	6	1.2	129	2	C84375	hypothetical prote	325	6	1.2	164	2	H65086	hypothetical prote
253	6	1.2	130	2	H95034	ribosomal protein	326	6	1.2	164	2	F91114	hypothetical prote
254	6	1.2	130	2	H97905	30S ribosomal prot	327	6	1.2	164	2	F85959	hypothetical prote
255	6	1.2	130	2	T37006	probable insertion	328	6	1.2	164	2	A10885	probable membrane
256	6	1.2	131	1	CCTW5T	cytochrome c552 (v	329	6	1.2	164	2	E87208	transcription elon
257	6	1.2	131	2	F64154	hypothetical prote	330	6	1.2	164	2	F70894	probable transcript
258	6	1.2	132	2	S73552	ribosomal protein	331	6	1.2	164	2	T23249	hypothetical prote
259	6	1.2	132	2	T12229	hypothetical prote	332	6	1.2	165	2	G72486	hypothetical prote
260	6	1.2	133	2	PH1287	Ig heavy chain pre	333	6	1.2	165	2	AE0110	probable membrane
261	6	1.2	133	2	F85917	hypothetical prote	334	6	1.2	165	2	TS2114	probable transcript
262	6	1.2	134	2	S00806	acyl carrier prote	335	6	1.2	166	2	T36616	hypothetical prote
263	6	1.2	134	2	S01257	acyl carrier prote	336	6	1.2	166	2	E97197	probable membrane
264	6	1.2	134	2	A26860	acyl carrier prote	337	6	1.2	166	2	B72607	hypothetical prote
265	6	1.2	134	2	S20499	acyl carrier prote	338	6	1.2	167	2	B64904	hypothetical prote
266	6	1.2	134	2	S10472	acyl carrier prote	339	6	1.2	167	2	D90892	fimbrial protein p
267	6	1.2	136	1	FOADH5	major core protein	340	6	1.2	167	2	E85725	probable fimbrial-
268	6	1.2	136	1	T09583	acyl carrier prote	341	6	1.2	167	2	G90175	NADH dehydrogenase
269	6	1.2	137	1	RSR128	ribosomal protein	342	6	1.2	168	2	T35737	probable integral
270	6	1.2	137	2	C84543	probable HsbB-like	343	6	1.2	169	2	T33727	conserved hypotet
271	6	1.2	137	2	AE0751	probable lipoprote	344	6	1.2	170	2	T48144	hypothetical prote
272	6	1.2	138	2	T39641	toxln 1 - Actinoba	345	6	1.2	171	2	B87593	hypothetical prote
273	6	1.2	138	2	S77160	acyl carrier prote	346	6	1.2	172	2	AF0311	NADH2 dehydrogenas
274	6	1.2	139	2	T01801	phocoystem I reac	347	6	1.2	172	2	C97894	hypothetical prote
275	6	1.2	139	2	AH1847	hypothetical prote	348	6	1.2	172	2	T088231	hypothetical prote
276	6	1.2	139	2	F97438	hypothetical prote	349	6	1.2	175	2	T13223	protein R175 - Lac
277	6	1.2	141	2	C69007	ribosomal protein	350	6	1.2	175	2	UC5912	Lyso protein - Lac
278	6	1.2	141	2	G82188	hypothetical prote	351	6	1.2	175	2	ACO148	auxin-induced prot
279	6	1.2	141	2	E70613	conserved hypotet	352	6	1.2	175	2	T02188	probable micrococ
280	6	1.2	142	2	E70613	probable ribosomal	353	6	1.2	175	2	H81372	ubiquinol-cytochro
281	6	1.2	142	2	AC3345	LSU ribosomal prot	354	6	1.2	176	2	A33382	hypothetical prote
282	6	1.2	142	2	A82517	ribonuclease P XP2	355	6	1.2	176	2	E90031	hypothetical prote
283	6	1.2	144	1	S52140	protein-tyrosine-p	356	6	1.2	176	2	F95347	protein imported
284	6	1.2	144	2	A83113	probable type II s	357	6	1.2	177	2	F72457	hypothetical prote
285	6	1.2	145	2	A53868	myoglobin I - L101	358	6	1.2	178	2	E75340	conserved hypotet
286	6	1.2	145	2	T29389	hypothetical prote	359	6	1.2	179	2	E75189	hypothetical prote
287	6	1.2	146	1	HEB02	hemoglobin epsilon	360	6	1.2	179	2	B72422	hypothetical prote
288	6	1.2	146	1	HEB04	hemoglobin epsilon	361	6	1.2	181	2	T79640	MHC cell surface a
289	6	1.2	147	1	HEG12	hemoglobin epsilon	362	6	1.2	181	2	C71662	hypothetical prote
290	6	1.2	147	1	F75182	ribosomal protein	363	6	1.2	181	2	H87384	acetyltransferase,
291	6	1.2	147	2	S18713	ribosomal protein	364	6	1.2	182	2	T51128	isopenentenyl diphos
292	6	1.2	147	2	B97174	chemotaxis protein	365	6	1.2	183	2	A48983	NADH2 dehydrogenas
293	6	1.2	147	2	AF1905	hypothetical prote	366	6	1.2	183	2	T46943	mannopine synthesi
294	6	1.2	147	2	H84341	hypothetical prote	367	6	1.2	184	2	S74232	bidirectional hydri
295	6	1.2	148	1	CKM6R	cytochrome c6 prec	368	6	1.2	184	2	D70362	ribosome recycling
296	6	1.2	148	2	A55712	hypothetical prote	369	6	1.2	184	2	T08200	probable disease r
297	6	1.2	148	2	E95384	protein imported	370	6	1.2	185	2	S59560	histone H1.41 - Ga
298	6	1.2	149	2	A47054	ferric uptake regu	371	6	1.2	185	2	D75635	resolvasase - Deinoc
299	6	1.2	149	2	A36870	ferric uptake regu	372	6	1.2	186	2	S04671	H+-transporting tw
300	6	1.2	150	2	S36292	T-cell receptor ga	373	6	1.2	186	2	B32806	resolvasase - Becher
301	6	1.2	150	2	D71031	probable ribosomal	374	6	1.2	186	2	S09632	tnpr protein - Pse
302	6	1.2	150	2	A42282	ferric uptake regu	375	6	1.2	186	4	S58115	resolvasase (EC 6.5
303	6	1.2	150	2	A82118	ferric uptake regu	376	6	1.2	187	2	T10073	amine dehydrogenas
304	6	1.2	150	2	T36099	hypothetical prote	377	6	1.2	187	2	C83627	alkyl hydroperoxid
305	6	1.2	150	2	E42528	B2AR protein - vac	378	6	1.2	188	2	B97875	conserved hypotet
306	6	1.2	150	2	S22208	hypothetical prote	379	6	1.2	189	2	A63024	conserved hypotet
307	6	1.2	150	2	A83919	hypothetical prote	380	6	1.2	189	2	F64533	co-chaperone and h
308	6	1.2	151	2	C70408	hypothetical prote	381	6	1.2	189	2	G72407	hypothetical prote
309	6	1.2	152	2	A95023	hypothetical prote	382	6	1.2	189	2	B95219	conserved hypotet
310	6	1.2	153	2	T00303	nuclease homolog -	383	6	1.2	189	2	H98082	hypothetical prote
311	6	1.2	153	2	S10665	hypothetical prote	384	6	1.2	190	2	B66871	hypothetical prote
312	6	1.2	153	2	AF1066	probable exported	385	6	1.2	191	2	H71973	24kDa chaperone -
313	6	1.2	155	2	T33008	hypothetical prote	386	6	1.2	193	2	H97732	calif thymus ribonu
314	6	1.2	156	2	B70077	hypothetical prote	387	6	1.2	194	2	G81437	endopeptidase Clp
315	6	1.2	159	2	AC3398	SSU ribosomal prot	388	6	1.2	195	2	S44788	D2007.2 protein -
316	6	1.2	159	2	H87539	conserved hypotet	389	6	1.2	196	1	HSXLSA	histone H5a - Afri
317	6	1.2	160	2	S74318	fimbrial protein c	390	6	1.2	196	1	ZZZRAM	nodulation protein
318	6	1.2	160	2	S74317	fimbrial protein c	391	6	1.2	196	1	ZZZRA4	nodulation protein
319	6	1.2	160	2	T35326	hypothetical prote	392	6	1.2	196	2	A30484	histone H5B - Afri
320	6	1.2	162	2	D83073	conserved hypotet	393	6	1.2	196	2	C95321	NodA N-acyltansfe
321	6	1.2	162	2	H70779	hypothetical prote	394	6	1.2	196	2	H87279	SCO1/2 family prot

395	6	1.2	197	2	AG3237	conserved hypochet	468	6	1.2	213	2	B97947	hypothetical prote
396	6	1.2	197	2	G82973	transcription regu	469	6	1.2	214	2	B87424	outer membrane pro
397	6	1.2	197	2	T38235	hypothetical prote	470	6	1.2	215	2	AH3389	transporter BME11
398	6	1.2	198	1	FOAD72	major core protein	471	6	1.2	215	2	T45229	probable F420H2-de
399	6	1.2	198	2	E86183	hypothetical prote	472	6	1.2	216	2	T12742	hypothetical prote
400	6	1.2	198	2	AC1880	hypothetical prote	473	6	1.2	218	2	A23055	histone H1.01 - ch
401	6	1.2	199	1	S47726	outer-membrane 1ip	474	6	1.2	218	2	AB3314	hydroxyacylglycath
402	6	1.2	199	2	I51280	atroxase (EC 3.4.2	475	6	1.2	219	2	B75440	hypothetical prote
403	6	1.2	199	2	A91176	outer membrane pro	476	6	1.2	219	2	B75440	rfbJ protein/conse
404	6	1.2	199	2	B86022	outer membrane pro	477	6	1.2	219	2	T38159	probable RNA-bindi
405	6	1.2	199	2	C71979	urease accessory p	478	6	1.2	220	2	AH3444	calf thymus ribonu
406	6	1.2	199	2	D64528	urease accessory p	479	6	1.2	220	2	E72204	hypothetical prote
407	6	1.2	199	2	T36905	hypothetical prote	480	6	1.2	223	1	F64404	hypothetical prote
408	6	1.2	200	1	A39741	cytochrome c bioge	481	6	1.2	223	2	UC6323	endonuclease (EC 3
409	6	1.2	200	2	AE1898	urease accessory p	482	6	1.2	223	2	T21437	hypothetical prote
410	6	1.2	200	2	A96664	unknown protein Tl	483	6	1.2	223	2	A40866	H1-60-induced diff
411	6	1.2	201	2	B87653	hypothetical prote	484	6	1.2	223	2	T30640	hypothetical prote
412	6	1.2	202	2	D81378	probable integral	485	6	1.2	223	2	E75390	phage shock protei
413	6	1.2	202	2	A82160	hypothetical prote	486	6	1.2	225	2	F87591	DNA-binding respon
414	6	1.2	202	2	S38922	hypothetical prote	487	6	1.2	225	2	E64075	urease accessory p
415	6	1.2	203	2	S46443	adamalysin (EC 3.4	488	6	1.2	225	2	*A83877	hypothetical prote
416	6	1.2	203	2	AB2871	urease accessory p	489	6	1.2	226	1	B64181	nrfB protein - Hae
417	6	1.2	203	2	S73281	hypothetical prote	490	6	1.2	227	2	T11327	cytochrome-c oxida
418	6	1.2	203	2	JC4871	phospholipase C (E	491	6	1.2	227	2	A72746	probable ribonucle
419	6	1.2	204	2	B45022	CRK-1 - human	492	6	1.2	227	2	A72746	amino acid ABC-cyp
420	6	1.2	204	2	AD3411	transporter, lyse	493	6	1.2	228	2	T38622	pyridoxal-phosphate
421	6	1.2	204	2	E72695	adenylate kinase (	494	6	1.2	228	2	AD0693	hypothetical prote
422	6	1.2	204	2	G83034	urease accessory p	495	6	1.2	228	2	A98166	hypothetical prote
423	6	1.2	204	2	F36950	urease accessory p	496	6	1.2	228	2	AF3121	hypothetical prote
424	6	1.2	204	2	H90027	urease accessory p	497	6	1.2	228	2	H97336	uncharacterized co
425	6	1.2	204	2	D97647	ureg protein (AB00	498	6	1.2	229	2	I51227	histone H1A - Afri
426	6	1.2	205	2	AG3014	Holliday junction	499	6	1.2	229	2	C82165	conserved hypochet
427	6	1.2	205	2	H98269	ruvA, holliday jun	500	6	1.2	229	2	C84248	galactosyltransfer
428	6	1.2	205	2	G90794	urease accessory p	501	6	1.2	230	2	E72276	phosphoribosylamin
429	6	1.2	205	2	T50714	urease accessory p	502	6	1.2	232	1	TVFV10	transforming prote
430	6	1.2	205	2	F36138	urease accessory p	503	6	1.2	232	2	JB0163	myelin expression
431	6	1.2	205	2	UN0755	urease accessory p	504	6	1.2	232	2	D90254	conserved hypochet
432	6	1.2	205	2	A85604	probable urease ac	505	6	1.2	233	2	UC7689	ubiquitin carboxyl
433	6	1.2	206	2	G82893	urease complex com	506	6	1.2	233	2	F64482	hypothetical prote
434	6	1.2	206	2	S74933	urease accessory p	507	6	1.2	233	2	F87216	probable TetR-famI
435	6	1.2	206	2	AC3478	urease accessory p	508	6	1.2	233	2	F72405	ABC transporter, A
436	6	1.2	207	2	A56190	protein ybJ's precu	509	6	1.2	234	2	B96875	probable membrane
437	6	1.2	207	2	E69085	titlin - rat (fragm	510	6	1.2	234	2	D70829	probable regulator
438	6	1.2	207	2	S54128	conserved hypochet	511	6	1.2	234	2	C75368	conserved hypochet
439	6	1.2	208	1	BVECB	incB protein - pha	512	6	1.2	235	2	B95889	probable transcrip
440	6	1.2	208	2	A83682	urease accessory p	513	6	1.2	235	2	G75324	hypothetical prote
441	6	1.2	208	2	AC3458	urease accessory p	514	6	1.2	236	2	B93083	hypothetical prote
442	6	1.2	208	2	T45272	methyltransferase	515	6	1.2	236	2	B69090	conserved hypochet
443	6	1.2	209	1	HSXLA	histone H1A - Afri	516	6	1.2	237	1	S56137	membrane-bound tet
444	6	1.2	209	2	S76990	hypothetical prote	517	6	1.2	237	2	F71563	probable sam-depen
445	6	1.2	209	2	AD2073	two-component resp	518	6	1.2	237	2	A81716	conserved hypochet
446	6	1.2	209	2	T40099	probable synaptoch	519	6	1.2	237	2	AB2623	hydrolyase (impo
447	6	1.2	209	2	G82359	conserved hypochet	520	6	1.2	237	2	AC1953	rubrerythrin (impo
448	6	1.2	210	2	T27471	hypothetical prote	521	6	1.2	238	2	B72320	ribonuclease HII -
449	6	1.2	211	2	G64136	probable DNA-tapur	522	6	1.2	238	2	S70468	agglutinin (WBA I)
450	6	1.2	211	2	C83809	uridine kinase udk	523	6	1.2	238	2	A55324	agglutinin (WBA I)
451	6	1.2	212	1	PMBYD	H+-transporting tw	524	6	1.2	238	2	D86538	SAM-dependent meth
452	6	1.2	212	2	A28470	histone H1 - mouse	525-	6	1.2	238	2	C72086	oxidereductase, sh
453	6	1.2	212	2	F64940	hypothetical prote	526	6	1.2	239	2	*G82383	blue fluorescent p
454	6	1.2	212	2	C90942	hypothetical prote	527	6	1.2	239	2	UC7768	epidermal growth f
455	6	1.2	212	2	G85790	hypothetical prote	528	6	1.2	239	2	A46243	fimbrial chaperone
456	6	1.2	212	2	AD0714	RNAase H (imported	529	6	1.2	239	2	AD0086	succinoglycan bios
457	6	1.2	212	2	A84349	hypothetical prote	530	6	1.2	241	2	H31844	KorA protein - Str
458	6	1.2	212	2	S25996	histone H1.3 - rab	531	6	1.2	241	2	F72285	conserved hypochet
459	6	1.2	213	1	HSRB13	GTP-binding protei	532	6	1.2	241	2	B87337	ada regulatory pro
460	6	1.2	213	2	S01766	GTP-binding protei	533	6	1.2	242	2	AB0165	arginine transport
461	6	1.2	213	2	F36364	GTP-binding protei	534	6	1.2	242	2	AH3521	high-affinity bran
462	6	1.2	213	2	S58279	hypothetical prote	535	6	1.2	243	2	C69473	methylthiodenosin
463	6	1.2	213	2	D69483	hypothetical prote	536	6	1.2	244	2	AB3524	transcription regu
464	6	1.2	213	2	F42365	hypothetical prote	537	6	1.2	245	2	B83462	probable short-cha
465	6	1.2	213	2	G93079	ABC transporter, A	538	6	1.2	246	2	C48492	3-deoxy-manno-octu
466	6	1.2	213	2	E90396	conserved hypochet	539	6	1.2	246	2	S46604	2-deoxyglucose-6-p
467	6	1.2	213	2	E83180	probable transcrip	540	6	1.2	246	2		

541	6	1.2	246	2	S46747	614	6	1.2	269	2	B71120	hypotheical prote
542	6	1.2	247	2	S77280	615	6	1.2	269	2	AD2833	GDEF family prote
543	6	1.2	247	2	S72859	616	6	1.2	270	2	T48780	hypotheical prote
544	6	1.2	247	2	G64788	617	6	1.2	270	2	B32835	hypotheical prote
545	6	1.2	248	1	J01682	618	6	1.2	273	2	AG1470	conserved hypochet
546	6	1.2	248	2	AG2969	619	6	1.2	274	2	J00686	nosinepride resist
547	6	1.2	248	2	G69111	620	6	1.2	274	2	D91178	probable ARAC-type
548	6	1.2	248	2	G97340	621	6	1.2	274	2	E86024	probable ARAC-type
549	6	1.2	249	2	E71649	622	6	1.2	274	2	S47736	probable transcrip
550	6	1.2	250	2	AG2635	623	6	1.2	274	2	AB1821	hypotheical prote
551	6	1.2	250	2	F97417	624	6	1.2	275	2	D98313	glutamine ABC tran
552	6	1.2	250	2	E82880	625	6	1.2	275	2	C84673	hypotheical prote
553	6	1.2	250	2	T08680	626	6	1.2	276	2	T45275	oxidoreductase of
554	6	1.2	251	2	C83053	627	6	1.2	276	2	T44386	ribosomal protein
555	6	1.2	251	2	C75521	628	6	1.2	276	2	B25345	tropoin T, cardia
556	6	1.2	252	2	A69096	629	6	1.2	276	2	A25345	tropoin T, cardia
557	6	1.2	252	2	B83750	630	6	1.2	276	2	T01177	hypotheical prote
558	6	1.2	253	1	PMHUTM	631	6	1.2	277	2	G73520	hypotheical prote
559	6	1.2	253	1	VHVUTV	632	6	1.2	278	2	B71896	probable outer mem
560	6	1.2	253	2	JM0040	633	6	1.2	278	2	H87663	hypotheical prote
561	6	1.2	253	2	AE2585	634	6	1.2	279	2	AG0421	sugar transport sy
562	6	1.2	253	2	AE2585	635	6	1.2	279	2	G81054	cytochrome c5 NM81
563	6	1.2	253	2	C82248	636	6	1.2	279	2	F81821	probable cytochrom
564	6	1.2	254	2	E71203	637	6	1.2	279	2	B97405	hydrolase, probabl
565	6	1.2	254	2	S10929	638	6	1.2	280	2	B87547	transcription regu
566	6	1.2	254	2	A49895	639	6	1.2	281	2	F82832	pantoate-beta-alan
567	6	1.2	254	2	D69140	640	6	1.2	282	2	A12156	S-formylglutathion
568	6	1.2	254	2	C75540	641	6	1.2	282	2	A87610	nicotinate-nucleot
569	6	1.2	255	2	D75415	642	6	1.2	282	2	B83453	hypotheical prote
570	6	1.2	255	2	A69359	643	6	1.2	282	2	F82851	conserved hypochet
571	6	1.2	255	2	AF3557	644	6	1.2	282	2	A97315	xylanase/chitin de
572	6	1.2	256	2	S14518	645	6	1.2	283	2	F87323	hypotheical prote
573	6	1.2	256	2	E72454	646	6	1.2	283	2	T12062	xpex protein - Xan
574	6	1.2	256	2	AH1978	647	6	1.2	283	2	T13879	matunase-like prot
575	6	1.2	257	2	T01342	648	6	1.2	283	2	H97610	hypotheical prote
576	6	1.2	257	2	E90763	649	6	1.2	284	1	S71851	heat shock transcr
577	6	1.2	257	2	F85626	650	6	1.2	284	2	H85436	heat shock transcr
578	6	1.2	258	2	E86786	651	6	1.2	284	2	G64753	YagM protein - Esc
579	6	1.2	258	2	A81448	652	6	1.2	284	2	T14167	ABC transport prot
580	6	1.2	258	2	G75372	653	6	1.2	284	2	H97991	hypotheical prote
581	6	1.2	258	2	T11582	654	6	1.2	285	2	D97367	moeb-like protein
582	6	1.2	259	2	A64509	655	6	1.2	285	2	UC4315	steroidogenic acut
583	6	1.2	259	2	A44988	656	6	1.2	285	2	G72414	hypotheical prote
584	6	1.2	259	2	H95133	657	6	1.2	285	2	T29804	hypotheical prote
585	6	1.2	259	2	B82348	658	6	1.2	286	2	A10288	pyridoxal kinase (
586	6	1.2	259	2	B96954	659	6	1.2	286	2	D71565	hypotheical prote
587	6	1.2	259	2	D98002	660	6	1.2	287	2	F64320	probable pyridoxal
588	6	1.2	259	2	E86288	661	6	1.2	287	2	A90922	pyridoxal kinase 2
589	6	1.2	260	2	B82927	662	6	1.2	287	2	E85770	pyridoxal kinase 2
590	6	1.2	260	2	F71438	663	6	1.2	288	2	E64151	probable pyridoxal
591	6	1.2	260	2	F95899	664	6	1.2	288	2	S69661	hypotheical prote
592	6	1.2	261	2	E83186	665	6	1.2	288	2	C82956	pyridoxamine kinas
593	6	1.2	262	2	T22130	666	6	1.2	288	2	PH1917	FL-160-1 protein -
594	6	1.2	262	2	AG0395	667	6	1.2	289	2	C84173	heat shock protein
595	6	1.2	264	1	ORBOLD	668	6	1.2	290	2	F90352	hypotheical prote
596	6	1.2	264	1	T14261	669	6	1.2	290	2	C96776	hypotheical prote
597	6	1.2	264	2	C87605	670	6	1.2	291	2	B95316	probable ABC trans
598	6	1.2	265	2	F86628	671	6	1.2	291	2	A81696	stationary-phase s
599	6	1.2	265	2	T36386	672	6	1.2	291	2	T20942	hypotheical prote
600	6	1.2	266	2	T37878	673	6	1.2	291	2	A98229	hypotheical membr
601	6	1.2	266	2	AC2383	674	6	1.2	291	2	H86075	hypotheical prote
602	6	1.2	267	1	TSPSNA	675	6	1.2	292	2	F71819	probable proteinas
603	6	1.2	267	1	JG6386	676	6	1.2	292	2	T09030	hypotheical prote
604	6	1.2	267	2	H96558	677	6	1.2	292	2	T35446	hypotheical prote
605	6	1.2	267	2	C72415	678	6	1.2	293	2	T04951	hypotheical prote
606	6	1.2	267	2	E83858	679	6	1.2	293	2	C90139	ribokinase (rbsK-1
607	6	1.2	268	1	A69000	680	6	1.2	293	2	E82582	DnaJ protein XP223
608	6	1.2	268	1	G83640	681	6	1.2	293	2	A71543	probable acid phos
609	6	1.2	269	1	J02127	682	6	1.2	293	2	H70625	hypotheical prote
610	6	1.2	269	1	A30768	683	6	1.2	293	2	E81661	conserved hypochet
611	6	1.2	269	2	A46506	684	6	1.2	294	2	AE1653	translation elonga
612	6	1.2	269	2	S28369	685	6	1.2	294	2	A11281	translation elonga
613	6	1.2	269	2	H69768	686	6	1.2	294	2	T35191	probable DNA hydro

687	6	1.2	296	2	A36366	enhancer-binding p	760	6	1.2	315	2	D84139	mannose-6-phosphat
688	6	1.2	296	2	F82077	nucleotide-nucleot	761	6	1.2	315	2	D85722	high-affinity bran
689	6	1.2	296	2	T51336	transcription acti	762	6	1.2	315	2	AC3522	probable membrane
690	6	1.2	296	2	A40593	transcription regu	763	6	1.2	315	2	E64906	conserved hypochet
691	6	1.2	296	2	T46828	heat shock protein	764	6	1.2	316	2	D70451	hypothetical prote
692	6	1.2	297	2	H82996	nicotinate-nucleot	765	6	1.2	316	2	D41830	hypothetical prote
693	6	1.2	297	2	E64733	quinolinate phosph	766	6	1.2	316	2	AG2977	probable transport
694	6	1.2	297	2	A99643	nicotinate-nucleot	767	6	1.2	316	2	C72615	hypothetical prote
695	6	1.2	297	2	A85494	quinolinate phosph	768	6	1.2	316	2	B70768	hypothetical prote
696	6	1.2	297	2	AH0530	nicotinate-nucleot	769	6	1.2	317	2	B82637	bifunctional biotri
697	6	1.2	298	2	B45470	hydroxymethylgluta	770	6	1.2	317	2	T38935	hypothetical prote
698	6	1.2	298	2	S75972	hypothetical prote	771	6	1.2	317	2	T10300	dUTP pyrophosphata
699	6	1.2	298	2	F84224	hypothetical prote	772	6	1.2	317	2	S63158	hypothetical prote
700	6	1.2	298	2	G90529	heat shock protein	773	6	1.2	318	2	C84651	hypothetical prote
701	6	1.2	298	2	A82879	conserved hypochet	774	6	1.2	319	2	F84966	chloredoxin-diulf
702	6	1.2	299	2	E69835	ribitol dehydrogen	775	6	1.2	319	2	E98305	hypothetical prote
703	6	1.2	299	2	H89906	hypothetical prote	776	6	1.2	319	2	T43040	hypothetical prote
704	6	1.2	300	2	S07540	site-specific DNA-	777	6	1.2	320	2	C81941	probable kinase/ph
705	6	1.2	300	2	S35549	site-specific DNA-	778	6	1.2	320	2	A81165	HPr kinase/phospha
706	6	1.2	300	2	D72405	hypothetical prote	779	6	1.2	320	2	A97279	hypothetical prote
707	6	1.2	301	1	RGECGT	GTP-binding protei	780	6	1.2	321	2	E87715	quinone oxidoreduc
708	6	1.2	301	1	J01927	nonstructural prot	781	6	1.2	321	2	AE3646	4-hydroxybutyrate
709	6	1.2	301	2	E81024	glycine-tRNA ligas	782	6	1.2	321	2	T08462	hypothetical prote
710	6	1.2	301	2	H91057	GTP-binding protei	783	6	1.2	322	2	T14597	proteinnase homolog
711	6	1.2	301	2	E85902	GTP-binding protei	784	6	1.2	322	2	AD3134	hypothetical prote
712	6	1.2	301	2	AB0829	GTP-binding protei	785	6	1.2	323	2	A99211	hypothetical prote
713	6	1.2	301	2	F90115	beta subunit of G	786	6	1.2	323	2	A72531	hypothetical prote
714	6	1.2	302	1	JN0794	phosphate butyrylc	787	6	1.2	324	2	A97493	hypothetical prote
715	6	1.2	302	1	C64103	glycine-tRNA ligas	788	6	1.2	324	2	AH2710	conserved hypochet
716	6	1.2	302	2	F64042	GTP-binding protei	789	6	1.2	324	2	D87075	probable DNA-bind
717	6	1.2	302	2	S72542	NSM protein - toma	790	6	1.2	325	1	H64056	protein-export pro
718	6	1.2	302	2	S52584	crte protein - Erw	791	6	1.2	325	2	A12252	adhesin precursor
719	6	1.2	302	2	A37802	crte protein - Erw	792	6	1.2	326	2	F75178	hydrogenase expers
720	6	1.2	302	2	AB0416	nicotinate-nucleot	793	6	1.2	326	2	E71035	probable hydrogena
721	6	1.2	302	2	D85025	hypothetical prote	794	6	1.2	326	2	F72639	hypothetical prote
722	6	1.2	303	1	A42544	nonstructural prot	795	6	1.2	327	2	B83995	branched-chain alp
723	6	1.2	303	2	AG0331	probable GTP-bind	796	6	1.2	327	2	S61982	probable membrane
724	6	1.2	303	2	S41754	CRKL protein - hum	797	6	1.2	327	2	AC2805	transporter Atu186
725	6	1.2	303	2	S58352	SH2/SH3 adaptor pr	798	6	1.2	327	2	C97584	hypothetical prote
726	6	1.2	303	2	G97734	methionyl-tRNA for	799	6	1.2	327	2	AB2135	conserved hypochet
727	6	1.2	303	2	F84401	hypothetical prote	800	6	1.2	327	2	AB8348	ferric ion ABC tra
728	6	1.2	304	2	A45022	CRK-II - human	801	6	1.2	328	2	D83312	hypothetical prote
729	6	1.2	304	2	E70698	C-Crk - mouse	802	6	1.2	328	2	T00747	RING-H2 finger pro
730	6	1.2	304	2	E70698	hypothetical prote	803	6	1.2	329	2	T35567	probable pantothen
731	6	1.2	305	2	B64981	hypothetical 32.6	804	6	1.2	329	2	F91290	hypothetical prote
732	6	1.2	305	2	B91006	probable transport	805	6	1.2	329	2	AB6132	hypothetical prote
733	6	1.2	305	2	C85850	probable transport	806	6	1.2	329	2	B75615	pyridoxamine kinas
734	6	1.2	305	2	AC0778	probable periplasm	807	6	1.2	329	2	H70744	hypothetical prote
735	6	1.2	306	2	D84039	hypothetical prote	808	6	1.2	329	2	T32272	hypothetical prote
736	6	1.2	306	2	F84276	formiminoglutamate	809	6	1.2	330	2	F97587	nicotinate-nucleot
737	6	1.2	306	2	D69298	conserved hypochet	810	6	1.2	331	2	T08403	photosystem II oxy
738	6	1.2	306	2	D70681	probable esterase/	811	6	1.2	332	1	B70015	chloredoxin reduct
739	6	1.2	306	2	T33487	hypothetical prote	812	6	1.2	332	2	AE0331	signal peptidase I
740	6	1.2	306	2	A43922	gravin - human (fr	813	6	1.2	332	2	E69148	UDP-glucose 4-epim
741	6	1.2	308	2	H82009	methionyl-tRNA for	814	6	1.2	332	2	C40835	kinesin-like prote
742	6	1.2	308	2	F81238	methionyl-tRNA for	815	6	1.2	332	2	C40646	endospore developm
743	6	1.2	308	2	T33676	hypothetical prote	816	6	1.2	332	2	T18527	anexin homolog -
744	6	1.2	309	2	AB0952	ribokinase (import	817	6	1.2	332	2	H98153	peptide ABC transp
745	6	1.2	309	2	AE3309	hypothetical prote	818	6	1.2	333	1	UC5166	ketol-acid reducto
746	6	1.2	310	2	S69050	probable membrane	819	6	1.2	333	2	D70855	probable ilvc prot
747	6	1.2	310	2	A55053	endothelial monocy	820	6	1.2	333	2	H87120	ketol-acid reducto
748	6	1.2	311	1	S02714	beta-lactamase (EC	821	6	1.2	333	2	B96657	probable replicati
749	6	1.2	311	2	T41284	hypothetical prote	822	6	1.2	333	2	H75458	probable low-affin
750	6	1.2	311	2	B84537	hypothetical prote	823	6	1.2	333	2	T36397	hypothetical prote
751	6	1.2	312	2	B69303	hypothetical prote	824	6	1.2	333	2	UC5050	sugar phosphate tr
752	6	1.2	312	2	T43853	conserved hypochet	825	6	1.2	333	2	UX0343	triacylglycerol li
753	6	1.2	312	2	S50430	hypothetical prote	826	6	1.2	334	2	S15941	PEP-fructosephosph
754	6	1.2	313	2	AD3208	dihydrodipicolinat	827	6	1.2	334	2	JU0298	Fructose repressor
755	6	1.2	314	2	T34369	hypothetical prote	828	6	1.2	334	2	D85490	Fructose repressor
756	6	1.2	314	2	AB3321	lysophospholipase	829	6	1.2	334	2	D90639	Fructose repressor
757	6	1.2	315	2	T09165	probable peroxidase	830	6	1.2	334	2	AC0517	fructose repressor
758	6	1.2	315	2	A90895	hypothetical prote	831	6	1.2	334	2	G85048	probable transposo
759	6	1.2	315	2	B96685	hypothetical prote	832	6	1.2	334	2	A12150	serine/chreonine k

833	6	1.2	335	2	T33483	hypothetical prote
834	6	1.2	335	2	A10481	pekb family carboh
835	6	1.2	336	2	A24430	glyceralddehyde-3-p
836	6	1.2	336	2	G69091	ribosomal protein
837	6	1.2	337	1	A23711	cytochrome-c oxida
838	6	1.2	337	2	S49554	hypothetical prote
839	6	1.2	338	1	C48648	ketol-acid reducto
840	6	1.2	338	2	AG2938	2-hydroxyacid-fam1
841	6	1.2	338	2	H98343	hypothetical prote
842	6	1.2	338	2	G70329	uroporphyrinogen d
843	6	1.2	338	2	C38163	nicotinate-nucleot
844	6	1.2	338	2	H75127	CAAX prenyl protei
845	6	1.2	339	2	AE3389	nicotinate-nucleot
846	6	1.2	339	2	E86393	protein T24P13.2 l
847	6	1.2	340	1	S04898	myb-related protei
848	6	1.2	340	2	E81331	uroporphyrinogen d
849	6	1.2	340	2	C55020	recombination prot
850	6	1.2	340	2	UX0292	recombination prot
851	6	1.2	340	2	B95147	alcohol dehydrogen
852	6	1.2	340	2	A35630	regulatory protein
853	6	1.2	340	2	G70432	conserved hypotet
854	6	1.2	341	1	RPECCT	transcription regu
855	6	1.2	341	2	E91236	transcription regu
856	6	1.2	341	2	E86083	transcription regu
857	6	1.2	341	2	AG0938	transcription regu
858	6	1.2	342	2	T16444	hypothetical prote
859	6	1.2	342	2	T36477	probable DNA-bind
860	6	1.2	342	2	T51703	nicotinate-nucleot
861	6	1.2	342	2	B87701	conserved hypotet
862	6	1.2	343	2	B83673	sorbitol dehydroge
863	6	1.2	343	2	T45415	ketol-acid reducto
864	6	1.2	343	2	D89605	protein F18G5.3 [1
865	6	1.2	343	2	S71094	rada protein - Hal
866	6	1.2	343	2	A81423	hypothetical prote
867	6	1.2	344	2	T05987	hypothetical prote
868	6	1.2	344	2	I49585	CD2 antigen protei
869	6	1.2	344	2	A69661	transcription regu
870	6	1.2	344	2	T38745	serine/chreonine p
871	6	1.2	344	2	AH2808	hypothetical prote
872	6	1.2	344	2	S19705	N5, N10-methylene
873	6	1.2	344	2	E95868	probable epoxide h
874	6	1.2	345	1	S58854	cell division prot
875	6	1.2	345	2	AG3186	hypothetical prote
876	6	1.2	345	2	T43736	transcription repr
877	6	1.2	345	2	AC1259	transcription repr
878	6	1.2	345	2	AG1621	transcription repr
879	6	1.2	346	2	F82349	ADP-heptose-Lbs he
880	6	1.2	346	2	T19676	hypothetical prote
881	6	1.2	346	2	AB3057	conserved hypotet
882	6	1.2	346	2	D98229	hypothetical prote
883	6	1.2	346	2	C98015	conserved hypotet
884	6	1.2	347	2	B69019	N5, N10-methylene
885	6	1.2	347	2	T35013	probable membrane
886	6	1.2	347	2	AH2950	conserved hypotet
887	6	1.2	347	2	G96700	protein F12A21.7 l
888	6	1.2	347	2	G84344	hypothetical prote
889	6	1.2	347	2	A81794	probable rotamase
890	6	1.2	347	2	H64371	malic acid transpo
891	6	1.2	347	2	S40733	hypothetical prote
892	6	1.2	347	2	AF2645	flagellar motor sw
893	6	1.2	347	2	F97427	flagellar motor sw
894	6	1.2	348	2	AG2043	RNA 3'-terminal ph
895	6	1.2	348	2	B81216	peptidyl-prolyl cl
896	6	1.2	348	2	T35968	conserved hypotet
897	6	1.2	348	2	T15219	hypothetical prote
898	6	1.2	348	2	AF3152	hypothetical prote
899	6	1.2	348	2	A12475	hypothetical prote
900	6	1.2	349	2	E75041	hypothetical prote
901	6	1.2	349	2	A71120	hypothetical prote
902	6	1.2	349	2	C72551	probable aminometh
903	6	1.2	350	2	B71473	probable Fe-S clus
904	6	1.2	350	2	S76698	hypothetical prote
905	6	1.2	350	2	T35849	probable ATP/GTP-b
906	6	1.2	350	2	E87327	hypothetical prote
907	6	1.2	351	2	C89788	sorbitol dehydroge
908	6	1.2	351	2	T26918	hypothetical prote
909	6	1.2	351	2	T44428	probable gonococca
910	6	1.2	351	2	AF0975	probable lacI fami
911	6	1.2	352	2	T49396	hypothetical prote
912	6	1.2	353	1	FOMWGR	gag polyprotein -
913	6	1.2	353	2	E86604	Fe-S oxidoreductas
914	6	1.2	353	2	F72019	conserved hypotet
915	6	1.2	354	2	D81735	conserved hypotet
916	6	1.2	354	2	E81281	hypothetical prote
917	6	1.2	354	2	A82544	translation initia
918	6	1.2	354	2	S60967	YGP1 protein precu
919	6	1.2	354	2	T24873	hypothetical prote
920	6	1.2	355	1	BVECMG	UDP-N-acetylglucos
921	6	1.2	355	2	F90640	hypothetical prote
922	6	1.2	355	2	B81777	UDP-N-acetylglucos
923	6	1.2	355	2	A81201	hypothetical prote
924	6	1.2	355	2	F85491	conserved hypotet
925	6	1.2	355	2	F69298	phcb protein (U601
926	6	1.2	356	2	AC0399	2-dehydro-3-deoxy-
927	6	1.2	356	2	A43680	L356 protein - Afr
928	6	1.2	356	2	I39510	carboxylesterase (
929	6	1.2	356	2	T51480	hypothetical prote
930	6	1.2	356	2	T37863	hypothetical sur1-
931	6	1.2	357	2	JO1647	SHL1 protein - hum
932	6	1.2	357	2	B83652	hypothetical prote
933	6	1.2	357	2	T12771	delta-endotoxin ho
934	6	1.2	357	2	AB3371	GTP-binding protei
935	6	1.2	357	2	WMBR38	infected cell prot
936	6	1.2	358	1	T44333	hypothetical prote
937	6	1.2	358	2	B87464	lysophospholipase
938	6	1.2	358	2	A83249	probable initiatio
939	6	1.2	358	2	H64397	N5, N10-methylene
940	6	1.2	358	2	F69809	spore germination
941	6	1.2	358	2	T00954	hypothetical prote
942	6	1.2	358	2	T48090	hypothetical prote
943	6	1.2	358	2	B40354	arylsulfatylphospha
944	6	1.2	359	2	T41906	hypothetical prote
945	6	1.2	359	2	H69329	nitrate ABC transp
946	6	1.2	359	2	E72290	branched chain am1
947	6	1.2	359	2	S44738	Cytoc2.5 protein -
948	6	1.2	359	2	JC7280	cytokine receptor-
949	6	1.2	359	2	A95277	hypothetical prote
950	6	1.2	359	2	G64557	GTP-binding protei
951	6	1.2	360	1	F64601	conserved hypotet
952	6	1.2	360	1	E70816	probable moa2 pro
953	6	1.2	360	2	B71952	hypothetical prote
954	6	1.2	360	2	E71910	hypothetical prote
955	6	1.2	360	2	C72263	motility protein p
956	6	1.2	361	2	JQ0374	lignin peroxidase
957	6	1.2	361	2	H70368	conserved phosphor
958	6	1.2	361	2	H70785	hypothetical prote
959	6	1.2	361	2	T30402	signal transductio
960	6	1.2	361	2	B96978	hypothetical prote
961	6	1.2	361	2	A96261	efflux protein [lm
962	6	1.2	361	2	AH3023	glycerol dehydroge
963	6	1.2	362	1	E64793	conserved hypotet
964	6	1.2	362	2	F82096	conserved esterase
965	6	1.2	363	2	T36408	hypothetical prote
966	6	1.2	363	2	T44150	fructose-bisphosph
967	6	1.2	363	1	ADHUB	hypothetical prote
968	6	1.2	364	1	C84076	hypothetical prote
969	6	1.2	364	2	C70777	probable cobC - My
970	6	1.2	364	2	T47198	H+-exporting ATPas
971	6	1.2	364	2	F90672	probable adhesin (
972	6	1.2	365	2	C69005	succinyl-CoA synth
973	6	1.2	365	2	JC2559	flagellin fljC-1 -
974	6	1.2	365	2	C86746	peptide chain rele
975	6	1.2	365	2	B69114	conserved hypotet
976	6	1.2	365	2	B72458	hypothetical prote
977	6	1.2	365	2	T50566	probable ABC-type
978	6	1.2	365	2		

```

979 6 1.2 365 2 S50409 hypothetical prote
980 6 1.2 366 2 S66016 probable GTP-bind
981 6 1.2 366 2 F84579 probable arginine
982 6 1.2 367 2 C89875 hypothetical prote
983 6 1.2 367 2 T26033 hypothetical prote
984 6 1.2 368 2 C95860 probable alcohol d
985 6 1.2 368 2 T14240 NADH2 dehydrogenas
986 6 1.2 368 2 G70370 conserved hypocher
987 6 1.2 368 2 A41111 flagellum-associat
988 6 1.2 368 2 AB2215 hypothetical prote
989 6 1.2 369 2 E86554 Fe-S oxidoreductas
990 6 1.2 369 2 G72069 conserved hypocher
991 6 1.2 369 2 B71516 probable Fe-S oxid
992 6 1.2 369 2 F81674 conserved hypocher
993 6 1.2 369 2 H82223 nicotinate-nucleot
994 6 1.2 369 2 B85523 hypothetical prote
995 6 1.2 370 1 D22930 DNA repair and gen
996 6 1.2 370 2 D83818 heat-shock protein
997 6 1.2 370 2 G84042 hypothetical prote
998 6 1.2 370 2 H71624 rifin PFB0030c - m
999 6 1.2 370 2 S49583 transcription fact
1000 6 1.2 371 2 D70357 conserved hypocher

```

## ALIGNMENTS

## RESULT 1

```

B81816 nitrite reductase (EC 1.7.99.3) NMA1887 [similarity] - Neisseria meningitidis (strain Z2
C/Species: Neisseria meningitidis
C/Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
R/Accession: B81816
R/Parikh, J.; Achman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel
; Holroyd, S.; Jørgensen, K.; Leather, S.; Moutre, S.; Mungall, K.; Quail, M.A.; Rajandream,
Nacure 404, 502-506, 2000
A/Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.
A/Reference number: A81775; MUID:20222556; PMID:10761919
A/Accession: B81816
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-386 <PAR>
A/Cross-references: GB:AL162757; GB:AL157959; NID:97380371; PIDN:CA85110.1; PID:9738052
A/Experimental source: serogroup A, strain Z2491
C/Genetics:
A:Gene: antiA; NMA1887
C/Keywords: oxidoreductase
F/330,171,179,184/Binding site: copper (His, Cys, His, Met) (type 1) #status predicted
F/335,170/Binding site: copper (His) (type 2) (shared with trimeric partner 2) #status p
F/325/Binding site: copper (His) (type 2) (shared with trimeric partner 1) #status predi

```

## Query Match

```

Best Local Similarity 6.8%; Score 34; DB 2; Length 386;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 178 YHCAVAPVGMHIANGMVGLIVPEKGLPRVDKE 211
Db 169 YHCAVAPVGMHIANGMVGLIVPEKGLPRVDKE 202

```

## RESULT 2

```

E81062 nitrite reductase (EC 1.7.99.3) NMB1623 [similarity] - Neisseria meningitidis (strain MC
N/Alternate names: outer membrane protein Pan1 homolog
C/Species: Neisseria meningitidis
C/Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C/Accession: E81062
R/Reichlin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;
ri, H.; Ok, H.; Vamathavan, J.; Gill, J.; Scarlato, V.; Maignani, V.; Pizza, M.
Science 287, 1809-1815, 2000
A/Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ve
A/Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.

```

```

A/Reference number: A81000; MUID:20175755; PMID:10710307
A/Accession: E81062
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-390 <TET>
A/Cross-references: GB:AE002512; GB:AE002098; NID:97226866; PIDN:AAFA1975.1; PID:972268
A/Experimental source: serogroup B, strain MC58
C/Genetics:
A:Gene: NMB1623
C/Keywords: oxidoreductase

```

## Query Match

```

Best Local Similarity 6.8%; Score 34; DB 2; Length 390;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 178 YHCAVAPVGMHIANGMVGLIVPEKGLPRVDKE 211
Db 173 YHCAVAPVGMHIANGMVGLIVPEKGLPRVDKE 206

```

## RESULT 3

```

A49208 nitrite reductase (EC 1.7.99.3) Pan1 [similarity] - Neisseria gonorrhoeae
N/Alternate names: outer membrane protein Pan1
C/Species: Neisseria gonorrhoeae
C/Date: 21-Jan-1994 #sequence_revision 18-Nov-1994 #text_change 15-Sep-2000
C/Accession: A49208
R/Hoehn, G.T.; Clark, V.L.
Infect. Immun. 60, 4704-4708, 1992
A/Title: The major anaerobically induced outer membrane protein of Neisseria gonorrhoea
A/Reference number: A49208; MUID:93014188; PMID:1198981
A/Accession: A49208
A/Status: preliminary; not compared with conceptual translation
A/Molecule type: DNA
A/Residues: 1-392 <HOE>
A/Note: sequence extracted from NCBI backbone (NCBIP:116468)
C/Keywords: oxidoreductase

```

## Query Match

```

Best Local Similarity 6.8%; Score 34; DB 2; Length 392;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 178 YHCAVAPVGMHIANGMVGLIVPEKGLPRVDKE 211
Db 173 YHCAVAPVGMHIANGMVGLIVPEKGLPRVDKE 206

```

## RESULT 4

```

S32112 (EC 1.7.2.1) (EC 1.7.2.1) - Pseudomonas aureofaciens
C/Species: Pseudomonas aureofaciens
C/Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 03-Jun-2002
C/Accession: S32112
R/Zumft, W.G.
Submitted to the EMBL Data Library, March 1993
A/Reference number: S32112
A/Accession: S32112
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-363 <ZUM>
A/Cross-references: EMBL:Z21945; NID:9287906; PIDN:CAA79939.1; PID:9287907
C/Keywords: copper; oxidoreductase

```

## Query Match

```

Best Local Similarity 2.4%; Score 12; DB 2; Length 363;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 137 MPMNVDFHATG 148
Db 111 MPMNVDFHATG 122

```

## RESULT 5



D64744  
exopolysaccharide synthetase regulator rcsf - Escherichia coli (strain K-12)  
C:Species: Escherichia coli  
C:Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 01-Mar-2002  
C:Accession: D64744; B47040  
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co  
S.; Rose, D.J.; Mau, B.; Shao, Y.  
Science 277, 1453-1462, 1997  
A:Title: The complete genome sequence of Escherichia coli K-12.  
A:Reference number: A64720; PMID:97426617; PMID:92785503  
A:Accession: D64744  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-134 <BLAT>  
A:Cross-references: GB:AE000128; GB:U00096; NID:g1786383; PIDN:AACT3307.1; PID:g1786394;  
R:Experimental source: strain K-12, substrain MG1655  
J:Geraiss, F.G.; Drapeau, G.R.  
J: Bacteriol. 174, 8016-8022, 1992  
A:Title: Identification, cloning, and characterization of rcsf, a new regulator gene for  
A:Reference number: A47040; PMID:93094132; PMID:1459951  
A:Accession: B47040  
A:Status: preliminary  
A:Molecule type: nucleic acid  
A:Residues: 1-36, DA, 39-45, 47-134 <GER>  
A:Cross-references: GB:L04474; NID:g147530; PIDN:AA24508.1; PID:g147532  
A:Note: sequence extracted from NCBI backbone (NCBIN:119949, NCBI:P:119951)  
C:Genetics:  
A:Gene: rcsf  
C:Function:  
A:Description: stimulates colanic acid capsule synthesis  
C:Superfamily: Escherichia coli exopolysaccharide synthetase regulator rcsf

Query Match  
Best Local Similarity 1.6%; Score 8; DB 2; Length 134;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 ALMLSGCS 24  
DB 10 ALMLSGCS 17

RESULT 6  
F90653  
regulator in colanic acid synthesis Rcsf [imported] - Escherichia coli (strain O157:H7,  
C:Species: Escherichia coli  
C:Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 03-Aug-2001  
C:Accession: F90653  
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.  
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
DNA Res. 8, 11-22, 2001  
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gene  
A:Reference number: A99629; PMID:21156231; PMID:11236796  
A:Accession: F90653  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-134 <HAY>  
A:Cross-references: GB:BA000007; PIDN:BA33621.1; PID:g13359654; GSPDB:GN00154  
C:Genetics:  
A:Gene: ECE098  
C:Superfamily: Escherichia coli exopolysaccharide synthetase regulator rcsf

Query Match  
Best Local Similarity 1.6%; Score 8; DB 2; Length 134;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 ALMLSGCS 24  
DB 10 ALMLSGCS 17

RESULT 7

F85504  
regulator in colanic acid synthesis Rcsf [imported] - Escherichia coli (strain O157:H7,  
C:Species: Escherichia coli  
C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 27-Nov-2001  
C:Accession: F85504  
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayh  
iller, L.; Grobbeck, E.J.; Davis, N.W.; Lm, A.; Dimianta, E.; Potamoukis, K.; Apodaca  
Nature 409, 529-533, 2001  
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
A:Reference number: A85480; PMID:21074935; PMID:11206551  
A:Accession: F85504  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-134 <STO>  
A:Cross-references: GB:AE005174; NID:g12512927; PIDN:AG54498.1; GSPDB:GN00145; UWGP:20  
A:Experimental source: strain O157:H7, substrain EDL933  
C:Genetics:  
A:Gene: rcsf  
C:Superfamily: Escherichia coli exopolysaccharide synthetase regulator rcsf

Query Match  
Best Local Similarity 1.6%; Score 8; DB 2; Length 134;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 ALMLSGCS 24  
DB 10 ALMLSGCS 17

RESULT 8  
J80215  
nitrite reductase (EC 1.7.99.3) blue copper-containing - Achromobacter xylosoxidans  
N:Alternate names: Nir  
C:Species: Achromobacter xylosoxidans  
C:Date: 21-Aug-1998 #sequence\_revision 21-Aug-1998 #text\_change 07-May-1999  
C:Accession: J80215  
R:Vandenberghe, I.H.M.; Meyer, T.E.; Cusanovich, M.A.; Van Beeumen, J.J.  
Biochem. Biophys. Res. Commun. 247, 734-740, 1998  
A:Title: The covalent structure of the blue copper-containing nitrite reductase from Aci  
A:Reference number: J80215; PMID:98321197; PMID:9647763  
A:Accession: J80215  
A:Molecule type: protein  
A:Residues: 1-336 <VAN>  
C:Comment: This enzyme is responsible for the conversion of nitrite to the gaseous prod  
C:Keywords: oxidoreductase

Query Match  
Best Local Similarity 1.6%; Score 8; DB 2; Length 336;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 137 MPNVDFF 144  
DB 87 MPNVDFF 94

RESULT 9  
JG0170  
nitrite reductase (EC 1.7.99.3) - Alcaligenes denitrificans subsp. xylosoxydans  
C:Species: Alcaligenes denitrificans subsp. xylosoxydans  
C:Date: 23-Jul-1999 #sequence\_revision 23-Jul-1999 #text\_change 11-May-2000  
C:Accession: JG0170  
R:Suzuki, E.; Horikoshi, N.; Kohzuma, T.  
Biochem. Biophys. Res. Commun. 255, 427-431, 1999  
A:Title: Cloning, sequencing, and transcriptional studies of the gene encoding copper-c  
A:Reference number: JG0170; PMID:99160880; PMID:10049725  
A:Accession: JG0170  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-360 <SUZ>  
A:Cross-references: GB:AF051831  
C:Keywords: copper; oxidoreductase

Query Match  
Best Local Similarity 1.6%; Score 8; DB 2; Length 360;

Best Local Similarity 100.0%; Pred. No. 13;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 137 MPNNVDFH 144  
DB 111 MPNNVDFH 118

## RESULT 10

D64333  
pyruvate synthase (EC 1.2.7.1) alpha chain - Methanococcus jannaschii

C/Species: Methanococcus jannaschii

C/Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 21-Jul-2000

C/Accession: D64333

R/Built: C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake, R.; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weissbrock, K.G.; Merrick, J.M.; Glodok, A.;

Science 273, 1058-1073, 1996

A/Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.

A/Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii

A/Reference number: A64300; MUID:96337999; PMID:8688087

A/Accession: D64333

A/Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-389 <BUL>

A/Cross-references: GB:U67482; GB:L77117; NID:g2826267; PIDN:AA98254.1; PID:g1590995; T

C/Genetics:

A/Map position: REV255064-253895

A/Start codon: GTG

C/Superfamily: pyruvate synthase alpha chain; 2-oxoacid ferredoxin oxidoreductase homolo

C/Keywords: coenzyme A; oxidoreductase

Query Match 1.6%; Score 8; DB 2; Length 389;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 340 FNKGALGI 347  
DB 332 FNKGALGI 339

## RESULT 11

F81221

hypothetical protein NMB0240 [imported] - Neisseria meningitidis (strain MCS8 serogroup

C/Species: Neisseria meningitidis

C/Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 19-Jan-2001

C/Accession: F81221

R/Built: H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.

Hickey, E.K.; Halt, D.R.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;

rti, H.; Qiu, H.; Vamathavan, J.; Gill, J.; Scariato, V.; Maignani, V.; Pizza, M.

Science 287, 1809-1815, 2000

A/Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappapoli, R.; Ve

A/Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MCS8.

A/Reference number: A81000; MUID:20175755; PMID:10710307

A/Accession: F81221

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-483 <TET>

A/Cross-references: GB:AE002381; GB:AE002098; NID:g7225455; PIDN:AAF40694.1; PID:g722546

C/Genetics:

A/Experimental source: serogroup B, strain MCS8

C/Superfamily: Neisseria meningitidis probable integral membrane protein NMA0020

Query Match 1.6%; Score 8; DB 2; Length 483;  
Best Local Similarity 100.0%; Pred. No. 16;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 35 SSTVDAAA 42  
DB 440 SSTVDAAA 447

## RESULT 12

H81992

probable integral membrane protein NMA0020 [imported] - Neisseria meningitidis (strain

C/Species: Neisseria meningitidis

C/Date: 05-May-2000 #sequence\_revision 05-May-2000 #text\_change 02-Feb-2001

C/Accession: H81992

R/Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel

J.; Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel

Nature 404, 502-506, 2000

A/Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.

A/Reference number: A81775; MUID:20222556; PMID:10761919

A/Accession: H81992

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-494 <PAR>

A/Cross-references: GB:AL162752; GB:AL157959; NID:g7378778; PIDN:CA83340.1; PID:g73787

A/Experimental source: serogroup A, strain Z2491

C/Genetics:

A/Superfamily: Neisseria meningitidis probable integral membrane protein NMA0020

Query Match 1.6%; Score 8; DB 2; Length 494;  
Best Local Similarity 100.0%; Pred. No. 17;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 35 SSTVDAAA 42  
DB 451 SSTVDAAA 458

## RESULT 13

F69802  
ABC transporter (ATP-binding protein) homolog yfiC - Bacillus subtilis

C/Species: Bacillus subtilis

C/Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 02-Feb-2001

C/Accession: F69802

R/Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Barte

C.; Bron, S.; Brouillet, S.; Bruscia, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch

A.; Ehrlich, S.D.; Emerson, P.T.; Eutian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.

Nature 390, 249-256, 1997

A/Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galle

lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.

Koetler, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,

A/Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mauee

Y, M.; Ogawa, K.; Ogiwara, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon

Rieger, M.; Schleich, S.; Schroeter, R.; Scoffone, P.; Sekiguchi, J.; Sekowska, A.; Sero

akeuchi, M.; Yamakoshi, A.; Yanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama

T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, I.

A/Authors: Yoshikawa, H.F.; Zumschein, E.; Yoshikawa, H.; Danchin, A.

A/Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.

A/Reference number: A69580; MUID:98044033; PMID:9384377

A/Accession: F69802

A/Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-604 <XUN>

A/Cross-references: GB:Z99108; GB:AL009126; NID:g2633055; PIDN:CA812651.1; PID:g2633146

A/Experimental source: strain 168

C/Genetics:

A/Gene: yfiC

C/Superfamily: Escherichia coli ABC transporter mlaA; ATP-binding cassette homology

C/Keywords: ATP; nucleotide binding; P-loop

F/382-576/Domain: ATP-binding cassette homology <ABC>

F/399-406/Region: nucleotide-binding motif A (P-loop)

Query Match 1.6%; Score 8; DB 2; Length 604;  
Best Local Similarity 100.0%; Pred. No. 20;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 40 AAKTANA 47  
DB 474 AAKTANA 481

## RESULT 14

T48418  
ABC transporter-like protein - Arabidopsis thaliana  
N/Alternate names: protein P8F6.120  
C/Species: Arabidopsis thaliana (mouse-ear cress)  
C/Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 20-Apr-2000  
A/Accession: T48418  
R/Bevan, M.; Peters, S.A.; van Staveren, M.; Dirkse, W.; Stiekema, W.; Bancroft, I.; New  
submitted to the Protein Sequence Database, March 2000  
A/Reference number: Z24488  
A/Accession: T48418  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-634 <BEV>  
A/Cross-references: EMBL:AL162873  
A/Experimental source: cultivar Columbia; BAC clone P8F6  
C/Genetics:  
A/Map position: 5  
A/Intons: 197/3; 226/3; 262/3; 351/3; 379/3; 452/1; 480/3; 543/2; 586/3  
A/Note: P8F6.120

Query Match  
Best Local Similarity 1.6%; Score 8; DB 2; Length 634;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 AAKTANAD 48  
|||||  
DB 508 AAKTANAD 515

RESULT 15  
T21147  
hypothetical protein F20D1.6 - Caenorhabditis elegans

C/Species: Caenorhabditis elegans  
C/Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 18-Feb-2000  
A/Accession: T21147  
R/Burton, J.  
submitted to the EMBL Data Library, August 1996  
A/Reference number: Z19382  
A/Accession: T21147  
A/Status: preliminary; translated from GB/EMBL/DDBJ  
A/Molecule type: DNA  
A/Residues: 1-915 <WIL>  
A/Cross-references: EMBL:Z78542; PIDN:CAB01748.1; GSPDB:GN00028; CESP:F20D1.6  
A/Experimental source: clone F20D1  
C/Genetics:  
A/Map position: X  
A/Intons: 45/3; 75/3; 126/3; 166/2; 196/3; 243/3; 271/2; 321/1; 388/3; 508/2; 552/1; 70  
C/Superfamily: Caenorhabditis elegans hypothetical protein F20D1.6

Query Match  
Best Local Similarity 1.6%; Score 8; DB 2; Length 915;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 87 KMETVEKV 94  
|||||  
DB 503 KMETVEKV 510

## RESULT 16

CCCF55  
cytochrome c55 - Chlorobium sp.  
C/Species: Chlorobium sp.  
C/Date: 13-Jul-1981 #sequence\_revision 13-Jul-1981 #text\_change 03-Mar-2000  
A/Accession: A00116  
R/Van Beeumen, J.; Ambler, R.P.; Meyer, T.E.; Kamen, M.D.; Olson, J.M.; Shaw, E.K.  
Biochem. J. 159, 757-774, 1976  
A/Title: The amino acid sequences of the cytochromes C-555 from two green sulphur bacteria  
A/Reference number: A00116; MUID:77087086; PMID:186412  
A/Accession: A00116  
A/Molecule type: protein

A/Residues: 1-86 <VAN>  
A/Note: the source is designated as Chlorobium thiosulfatophilum  
R/Korszun, Z.R.; Saleme, F.R.  
Proc. Natl. Acad. Sci. U.S.A. 74, 5244-5247, 1977

A/Title: Structure of cytochrome c55 of Chlorobium thiosulfatophilum: primitive low-potential  
A/Reference number: A38043; MUID:78094383; PMID:282997  
A/Content: annotation; X-ray crystallography, 2.7 angstroms  
C/Comment: This basic C-type monoheme cytochrome has been found exclusively in the green  
low redox potential compared with mitochondrial cytochrome c. It is reactive with cyto-  
C/Superfamily: cytochrome c6; cytochrome c6 homology  
C/Keywords: chromoprotein; electron transfer; heme; iron; metalloprotein; photosynthesis  
P/4-81/Domain: cytochrome c6 homology <CYC>  
P/14,17/Binding site: heme (Cys) (covalent) #status experimental  
P/18,60/Binding site: heme iron (His, Met) (axial ligands) #status experimental

Query Match  
Best Local Similarity 1.4%; Score 7; DB 1; Length 86;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 397 AGKATYD 403  
|||||  
DB 5 AGKATYD 11

## RESULT 17

AG3064  
hypothetical protein Atcu4133 [imported] - Agrobacterium tumefaciens (strain C58, Dupont)  
C/Species: Agrobacterium tumefaciens  
C/Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 18-Nov-2002  
A/Accession: AG3064  
R/Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.  
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClellan,  
; Karp, P.; Romero, P.; Zhang, S.  
Science 294, 2317-2323, 2001  
A/Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,  
ster, E.W.  
A/Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
A/Reference number: AB2577; MUID:21608550; PMID:11743193  
A/Accession: AG3064  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-98 <KUR>  
A/Cross-references: GB:AE008669; PIDN:AAL44933.1; PID:g17742587; GSPDB:GN00187  
A/Experimental source: strain C58 (Dupont)  
C/Genetics:  
A/Map position: linear chromosome

Query Match  
Best Local Similarity 1.4%; Score 7; DB 2; Length 98;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 270 RLTVGNG 276  
|||||  
DB 92 RLTVGNG 98

## RESULT 18

S52930  
GP41 ENV protein - human immunodeficiency virus type 1 (fragment)  
C/Species: human immunodeficiency virus type 1, HIV-1  
C/Date: 06-Jun-1995 #sequence\_revision 21-Jul-1995 #text\_change 26-Aug-1999  
A/Accession: S52930  
R/Cohen, J.H.M.; Guetard, D.; Philbert, F.; Chamaret, S.; Taly, T.; Montagnier, L.; d  
submitted to the EMBL Data Library, January 1995  
A/Description: A novel HIV-1 O strain illustrates the diversity of the O group.  
A/Reference number: S52929  
A/Accession: S52930  
A/Status: preliminary  
A/Molecule type: genomic RNA  
A/Residues: 1-104 <COH>  
A/Cross-references: EMBL:X84328; NID:g695526; PIDN:CAAS9066.1; PID:g695527  
C/Superfamily: type E retrovirus env polypeptide

Query Match 1.4%; Score 7; DB 2; Length 104;  
Best Local Similarity 100.0%; Pred. No. 46;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 467 DQDIANY 473  
|||||  
DB 83 DQDIANY 89

## RESULT 19

ribulose-bisphosphate carboxylase (EC 4.1.1.39) small chain - Prochlorothrix hollandica  
C/Species: Prochlorothrix hollandica  
C/Date: 21-Nov-1993 #sequence\_revision 10-Nov-1995 #text\_change 22-Jun-1999  
C/Accession: S16437

R/Morden, C.W.; Golden, S.S.  
J. Mol. Evol. 32, 379-395, 1991  
A/Title: Sequence analysis and phylogenetic reconstruction of the genes encoding the large  
aroyte Prochlorothrix hollandica.  
A/Reference number: S16436; MUID:91251137; PMID:1904095  
A/Accession: S16437  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-109 <MOR>

A/Cross-references: EMBL:X57359; NID:945542; PIDN:CAA40633.1; PID:945544  
C/Superfamily: ribulose-bisphosphate carboxylase small chain  
C/Keywords: carbon-carbon lyase; carboxy-lyase

Query Match 1.4%; Score 7; DB 2; Length 109;  
Best Local Similarity 100.0%; Pred. No. 48;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 465 LSDQQA 471  
|||||  
DB 19 LSDQQA 25

## RESULT 20

hypothetical protein aq\_1239 - Aquifex aeolicus  
C/Species: Aquifex aeolicus  
C/Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 09-Jun-2000  
C/Accession: A70407

R/Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Oy  
V.  
Nature 392, 353-358, 1998

A/Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.  
A/Reference number: A70300; MUID:98196666; PMID:9537320  
A/Accession: A70407

A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
A/Molecule type: DNA  
A/Residues: 1-111 <AQP>  
A/Cross-references: GB:AE000730; NID:92983674; PIDN:AAC07249.1; PID:92983684; GB:AE00065  
A/Experimental source: strain VFS  
C/Genetics:  
A/Gene: aq\_1239  
C/Superfamily: Aquifex aeolicus hypothetical protein aq\_1239

Query Match 1.4%; Score 7; DB 2; Length 111;  
Best Local Similarity 100.0%; Pred. No. 49;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 228 QGLOPFD 234  
|||||  
DB 28 QGLOPFD 34

## RESULT 21

ribosomal protein S9 [similarity] - Lactococcus lactis subsp. lactis (strain IL1403)  
C/Species: Lactococcus lactis subsp. lactis  
C/Date: 23-Mar-2001 #sequence\_revision 23-Mar-2001 #text\_change 03-May-2002

C/Accession: E86906  
R/Bolotin, A.; Winkler, P.; Mager, S.; Jallion, O.; Malarme, K.; Weisenbach, J.; Ehrlich,  
Genome Res. 11, 731-753, 2001  
A/Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis s

A/Reference number: A86625; MUID:21235186; PMID:11337471  
A/Accession: E86906  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-130 <STO>  
A/Cross-references: GB:AE005176; PID:912725323; PIDN:AAK06351.1; GSPDB:GN00146  
A/Experimental source: strain IL1403  
C/Genetics:  
A/Gene: rps1  
C/Superfamily: Escherichia coli ribosomal protein S9

Query Match 1.4%; Score 7; DB 2; Length 130;  
Best Local Similarity 100.0%; Pred. No. 56;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 447 GKITVNG 453  
|||||  
DB 26 GKITVNG 32

## RESULT 22

transcription elongation factor grea [imported] - Brucella melitensis (strain 16M)  
C/Species: Brucella melitensis  
C/Date: 01-Feb-2002 #sequence\_revision 01-Feb-2002 #text\_change 15-Feb-2002  
C/Accession: AF3315

R/DiVincenzo, V.G.; Kapral, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova,  
.; Mazur, M.; Goltzman, E.; Selkov, E.; Elser, P.H.; Hagius, S.; O'Callaghan, D.; Letes  
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002  
A/Title: The genome sequence of the facultative intracellular pathogen Brucella meliten

A/Reference number: AD3252; PMID:11756688  
A/Accession: AF3315  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-157 <KUR>  
A/Cross-references: GB:AE008917; PIDN:AAL51689.1; PID:917982422; GSPDB:GN00190  
A/Experimental source: strain 16M  
C/Genetics:  
A/Gene: BMEI0508  
A/Map position: I

C/Superfamily: transcription elongation factor greb  
C/Keywords: transcription factor

Query Match 1.4%; Score 7; DB 2; Length 157;  
Best Local Similarity 100.0%; Pred. No. 65;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 121 EGDTEV 127  
|||||  
DB 134 EGDTEV 140

## RESULT 23

hypothetical protein RC0818 [imported] - Rickettsia conorii (strain Malish 7)  
C/Species: Rickettsia conorii  
C/Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 22-Oct-2001  
C/Accession: B97802

R/Ogata, H.; Audic, S.; Renesto-Audiffren, P.; Fournier, P.E.; Barbe, V.; Samson, D.; R  
Science 293, 2093-2098, 2001  
A/Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.  
A/Reference number: A97700; MUID:21442074; PMID:11557893  
A/Accession: B97802  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-161 <KUR>

A/Cross-references: GB:AE006914; PIDN:AAL03356.1; PID:915619918; GSPDB:GN00173  
C/Genetics:  
A/Gene: RC0818

C:Superfamily: nus operon 15K protein

Query Match 1.4%; Score 7; DB 2; Length 161;  
Best Local Similarity 100.0%; Pred. No. 67;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 468 QOIANYI 474  
|||||  
DB 6 QOIANYI 12

RESULT 24

A97742 cytochrome c [imported] - Rickettsia conorii (strain Malish 7)

C:Species: Rickettsia conorii

C>Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 09-Nov-2001

C:Accession: A97742

R:Ogata, H.; Audic, S.; Renesto-Audiffren, P.; Fournier, P.E.; Barde, V.; Samson, D.; Ro

Science 293, 2093-2098, 2001

A:Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.

A:Reference number: A97700; PMID:21442074; PMID:11557893

A:Accession: A97742

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-175 <KUR>

A:Cross-references: GB:AE006914; PIDN:AAL02875.1; PID:G15619399; GSPDB:GN00173

C:Genetics:

A:Gene: CYCM

C:Superfamily: membrane-bound cytochrome c/cym; cytochrome c homology

C:Keywords: chromoprotein; heme; iron; metalloprotein

F:84,87/Binding site: heme (Cys) (covalent) #status predicted

F:88/Binding site: heme iron (His) (axial ligand) #status predicted

F:150/Binding site: heme iron (Met) (axial ligand) #status predicted

Query Match 1.4%; Score 7; DB 2; Length 175;

Best Local Similarity 100.0%; Pred. No. 72;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 43 KTNANDN 49  
|||||  
DB 70 KTNANDN 76

RESULT 25

AH0196 probable lipoprotein YPO1613 [imported] - Yersinia pestis (strain CO92)

C:Species: Yersinia pestis

C>Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 02-Nov-2001

C:Accession: AH0196

R:Parhill, J.; Wren, B.W.; Thomson, N.R.; Tibball, R.W.; Holden, M.T.G.; Prentice, M.B.

demo-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Dougan, G.;

11, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrett,

Nature 413, 523-527, 2001

A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.

A:Reference number: AB0001; PMID:21470413; PMID:11586360

A:Accession: AH0196

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-191 <KUR>

A:Cross-references: GB:AL590842; PIDN:CAC90435.1; PID:G15979651; GSPDB:GN00175

C:Genetics:

A:Gene: YPO1613

Query Match 1.4%; Score 7; DB 2; Length 191;

Best Local Similarity 100.0%; Pred. No. 77;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 71 EYPPVD 77  
|||||  
DB 40 EYPPVD 46

RESULT 26

T29645 hypothetical protein W01A11.4 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 02-Jun-2000

C:Accession: T29645

R:Blanchard, M.; Bradshaw, H.

submitted to the EMBL Data Library, July 1996

A:Description: The sequence of C. elegans cosmid W01A11.

A:Reference number: Z20658

A:Accession: T29645

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-192 <BLA>

A:Cross-references: EMBL:U64852; PIDN:AB04967.1; GSPDB:GN00023; CESP:W01A11.4

A:Experimental source: strain Bristol N2; clone W01A11

C:Genetics:

A:Gene: W01A11.4

A:Map position: 5

A:Insertions: 7/3; 101/3

C:Superfamily: beta-galactoside-binding lectin

Query Match 1.4%; Score 7; DB 2; Length 192;  
Best Local Similarity 100.0%; Pred. No. 77;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 292 DKVHEFG 298  
|||||  
DB 133 DKVHEFG 139

RESULT 27

F64941 rnd protein - Escherichia coli (strain K-12)

C:Species: Escherichia coli

C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 01-Mar-2002

C:Accession: F64941; S41588

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; C

.A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; PMID:97426617; PMID:9278503

A:Accession: F64941

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-193 <BLAT>

A:Cross-references: GB:AE000275; GB:U00096; NID:G1788106; PIDN:ACT4876.1; PID:G1788108

A:Experimental source: Strain K-12, substrain MG1655

Mol. Gen. Genet. 242, 241-249, 1994

A:Title: The fadD gene of Escherichia coli K12 is located close to rnd at 39.6 min of t

A:Reference number: S41588; PMID:94150456; PMID:8107670

A:Accession: S41588

A:Molecule type: DNA

A:Residues: 121-123, 'P', 125-193 <FUD>

C:Genetics:

A:Gene: rnd

C:Superfamily: rnd protein

Query Match 1.4%; Score 7; DB 1; Length 193;

Best Local Similarity 100.0%; Pred. No. 78;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 ALMLSGC 23  
|||||  
DB 17 ALMLSGC 23

RESULT 28

C90943

probable outer membrane protein ECG2515 [imported] - Escherichia coli (strain O157:H7,

C:Species: Escherichia coli

C>Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 03-Aug-2001

C:Accession: C90943

R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;  
gasawara, N.; Yasunaga, T.; Kuhnara, S.; Shiba, T.; Hatori, M.; Shinagawa, H.  
DNA Res. 8, 11-22, 2001  
A>Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and gen  
A;Reference number: A96629; MUID:2156233; PMID:11258796  
A;Accession: C90943  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-193 <HAV>  
A;Cross-references: GB:BA000007; PIDN:BAH35938.1; PID:G13361982; GSPDB:GN00154  
A;Experimental source: strain O157:H7, substrain R1MD 0509552  
A;Gene: EC62515  
C;Superfamily: rnd protein

Query Match 1.4%; Score 7; DB 2; Length 193;  
Best Local Similarity 100.0%; Pred. No. 78;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 ALMLSGC 23  
Db 17 ALMLSGC 23

RESULT 29  
G85791  
Probable outer membrane protein Z2849 [imported] - *Escherichia coli* (strain O157:H7, sub  
C;Species: *Escherichia coli*  
C;Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001  
C;Accession: G85791  
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
Miller, L.; Grobocck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,  
Nature 409, 529-533, 2001  
A>Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.  
A;Reference number: A85480; MUID:21074935; PMID:11206551  
A;Accession: G85791  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-193 <STO>  
A;Cross-references: GB:AE005174; NID:G12515846; PIDN:AAG56795.1; GSPDB:GN00145; UMGF:Z28  
C;Experimental source: strain O157:H7, substrain EDL933  
A;Gene: Z2849  
C;Superfamily: rnd protein

Query Match 1.4%; Score 7; DB 2; Length 193;  
Best Local Similarity 100.0%; Pred. No. 78;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 ALMLSGC 23  
Db 17 ALMLSGC 23

RESULT 30  
AD0725  
Probable lipoprotein STY1949 [imported] - *Salmonella enterica* subsp. *enterica* serovar Ty  
C;Species: *Salmonella enterica* subsp. *enterica* serovar Typhi  
A;Note: this species has also been called *Salmonella typhi*  
C;Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 18-Nov-2002  
C;Accession: AD0725  
R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Main, J.; Churcher,  
th, T.; Conerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,  
S.; Moule, S.; O'Garra, P.  
Nature 413, 848-852, 2001  
A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;  
A>Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* serov  
A;Reference number: AB0502; MUID:21534947; PMID:11677608  
A;Accession: AD0725  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-193 <PAR>  
A;Cross-references: GB:AL513382; PIDN:CAD05502.1; PID:G16503006; GSPDB:GN00176

C;Genetics:  
A;Gene: STY1949  
C;Superfamily: rnd protein

Query Match 1.4%; Score 7; DB 2; Length 193;  
Best Local Similarity 100.0%; Pred. No. 78;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 ALMLSGC 23  
Db 17 ALMLSGC 23

RESULT 31  
G69539  
ribosomal protein L15, cytosolic - *Archaeoglobus fulgidus*  
C;Species: *Archaeoglobus fulgidus*  
C;Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 02-Feb-2001  
C;Accession: G69539  
R;Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson  
; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F  
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.  
Nature 390, 364-370, 1997  
A;Authors: Uitterlinden, T.; Cotton, M.D.; Spriggs, T.; Artach, P.; Kaime, B.P.; Sykes, S  
Smith, H.O.; Woese, C.R.; Venter, J.C.  
A>Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archae  
A;Reference number: A69250; MUID:98049343; PMID:9389475  
A;Accession: G69539  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-194 <LEB>  
A;Cross-references: GB:AE000944; GB:AE000782; NID:G2689267; PIDN:AAB88937.1; PID:G26482  
C;Keywords: cytosol; nucleotide binding; P-loop; protein biosynthesis; ribosome  
P160-168/Region: nucleotide-binding motif A (P-loop)

Query Match 1.4%; Score 7; DB 2; Length 194;  
Best Local Similarity 100.0%; Pred. No. 78;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 76 VDRDHPA 82  
Db 134 VDRDHPA 140

RESULT 32  
T36138  
hypothetical protein SCE19A.15c - *Streptomyces coelicolor*  
C;Species: *Streptomyces coelicolor*  
C;Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 08-Sep-2000  
C;Accession: T36138  
R;Seeger, K.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.  
submitted to the EMBL Data Library, July 1999  
A;Reference number: Z21598  
A;Accession: T36138  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-202 <SEB>  
A;Cross-references: EMBL:AL096857; PIDN:CAB50996.1; GSPDB:GN000070; SCOEDB:SCE19A.15c  
A;Experimental source: strain A13(2)  
C;Genetics:  
A;Gene: SCOEDB:SCE19A.15c  
C;Superfamily: Streptomyces coelicolor hypothetical protein SCE19A.15c

Query Match 1.4%; Score 7; DB 2; Length 202;  
Best Local Similarity 100.0%; Pred. No. 81;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 370 PEGAPOA 376  
Db 77 PEGAPOA 83

RESULT 33  
A:Gene: AT4G09590  
A:Residues: 1-212 <STO>  
A:Cross-references: GB:Z77165; GB:AU123456; NID:g3261609; PIDN:CAB01009.1; PID:e255083;  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C/Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 16-Feb-2001  
C/Accession: A85098  
R:Anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spring  
Nature 402, 769-777, 1999  
A/Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.  
A/Reference number: A85001; MUID:20083488; PMID:10617198  
A/Accession: A85098  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-211 <STO>  
A:Cross-references: GB:NC\_001268; NID:g7267654; PIDN:CAB78082.1; GSPDB:GN00140  
C/Genetics:  
A:Gene: AT4G09590  
A/Map position: 4

Query Match  
Best Local Similarity 1.4%; Score 7; DB 2; Length 211;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 NQADKAA 31  
DB 184 NQADKAA 190

RESULT 34  
C84775  
probable harpin-induced protein [imported] - Arabidopsis thaliana  
C/Species: Arabidopsis thaliana (mouse-ear cress)  
C/Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
C/Accession: C84775  
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Bent, M.T.; Town, C.D.; Fujii, C.Y.;  
M.; Koo, H.; Molnar, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Tallon, L.;  
Eus, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.  
Nature 402, 761-768, 1999  
A/Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A/Reference number: A84420; MUID:20083487; PMID:10617197  
A/Accession: C84775  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-211 <STO>  
A:Cross-references: GB:AE002093; NID:g4510372; PIDN:AAD21460.1; GSPDB:GN00139  
C/Genetics:  
A:Gene: At2g35970  
A/Map position: 2

Query Match  
Best Local Similarity 1.4%; Score 7; DB 2; Length 211;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 NQADKAA 31  
DB 184 NQADKAA 190

RESULT 35  
C70737  
probable sigd protein - Mycobacterium tuberculosis (strain H37RV)  
C/Species: Mycobacterium tuberculosis  
C/Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 22-Oct-1999  
C/Accession: C70737  
R:Cole, S.T.; Broesch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.;  
Conroy, R.; Davies, R.; Devlin, K.; Fellwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;  
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998  
A:Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Bartell, B.G.  
A/Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A/Reference number: A70500; MUID:98295987; PMID:9634230  
A/Accession: C70737  
A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA  
A:Residues: 1-212 <COL>  
A:Cross-references: GB:Z77165; GB:AU123456; NID:g3261609; PIDN:CAB01009.1; PID:e255083;  
A:Experimental source: strain H37RV  
C/Genetics:  
A:Gene: sigd

Query Match  
Best Local Similarity 1.4%; Score 7; DB 2; Length 212;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 488 LSADVVA 494  
DB 71 LSADVVA 77

RESULT 36  
T29906  
hypothetical protein F59A3.5 - Caenorhabditis elegans  
C/Species: Caenorhabditis elegans  
C/Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C/Accession: T29906  
R:Hu, X.; Le, T.T.  
Submitted to the EMBL Data Library, November 1996  
A/Description: The sequence of C. elegans cosmid F59A3.  
A/Reference number: Z20707  
A/Accession: T29906  
A>Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-212 <WUX>  
A:Cross-references: EMBL:U80448; PIDN:AAB37817.1; GSPDB:GN00019; CESP:F59A3.5  
A:Experimental source: strain Bristol N2; clone F59A3  
C/Genetics:  
A:Gene: CESP:F59A3.5  
A/Map position: 1  
A/Intons: 12/1; 43/3; 78/3; 108/3; 155/1

Query Match  
Best Local Similarity 1.4%; Score 7; DB 2; Length 212;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 260 ALKAKVG 266  
DB 104 ALKAKVG 110

RESULT 37  
S55925  
probable arabinogalactan protein precursor - tomato  
C/Species: Lycopersicon esculentum (tomato)  
C/Date: 28-Oct-1995 #sequence\_revision 03-Nov-1995 #text\_change 21-Jul-2000  
C/Accession: S55925  
R:Fogson, B.O.; Davies, C.  
Plant Mol. Biol. 28, 347-352, 1995  
A/Title: Characterization of a cDNA encoding the protein moiety of a putative arabinogal  
A/Reference number: S55925; MUID:95322597; PMID:7599320  
A/Accession: S55925  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-215 <POG>  
A:Cross-references: EMBL:247980; NID:g872126; PIDN:CAA8023.1; PID:g872127  
C/Superfamily: proline-rich protein

Query Match  
Best Local Similarity 1.4%; Score 7; DB 2; Length 215;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 383 PKTPAPA 389  
DB 45 PKTPAPA 51

RESULT 38

G91207  
 Probable replicase [imported] - Escherichia coli (strain O157:H7, substrain RMD 0509952)  
 C:Species: Escherichia coli  
 C>Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 17-May-2002  
 C:Accession: G91207  
 R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;  
 Ganeswara, N.; Yasunaga, T.; Kunihara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
 DNA Ref: 8, 11-22, 2001  
 A>Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gene  
 A:Reference number: A99629; MUID:2156231; PMID:11258796  
 A:Accession: G91207  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-218 <HAV>  
 A:Cross-references: GB:BA000007; PIDN:BA938054.1; PID:G13364106; GSPDB:GN00154  
 A:Experimental source: strain O157:H7, substrain RMD 0509952  
 C:Gene: Ec94631  
 C:Superfamily: Escherichia coli hypothetical protein yIdX

Query Match 1.4%; Score 7; DB 2; Length 218;  
 Best Local Similarity 100.0%; Pred. No. 86;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 ALMLSGC 23  
 |||||  
 Db 18 ALMLSGC 24

## RESULT 39

T15381  
 hypothetical protein CO3B1.9 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C>Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999  
 C:Accession: T15381  
 R:Martin, J.  
 submitted to the EMBL Data Library, November 1995  
 A:Description: The sequence of C. elegans cosmid CO3B1.  
 A:Reference number: Z18340  
 A:Accession: T15381  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-230 <MAR>  
 A:Cross-references: EMBL:U40952; NID:G1072237; PID:G1072239; PIDN:AAA81738.1; CESP:CO3B1  
 C:Gene: CESP:CO3B1.9  
 A:Introns: 29/1; 116/1; 134/3; 192/3

Query Match 1.4%; Score 7; DB 2; Length 230;  
 Best Local Similarity 100.0%; Pred. No. 90;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 449 ITVNGNQ 455  
 |||||  
 Db 58 ITVNGNQ 64

## RESULT 40

I38849  
 LERK-3 - human  
 C:Species: Homo sapiens (man)  
 C>Date: 29-May-1998 #sequence\_revision 29-May-1998 #text\_change 29-Sep-1999  
 C:Accession: I38849  
 R:Kozlovsky, C.J.; Maraskovsky, E.; McGrew, J.T.; VandenBos, T.; Teepe, M.; Lyman, S.D.;  
 Oncogene 10, 299-306, 1995  
 A>Title: Ligands for the receptor tyrosine kinases hek and elk: isolation of cDNAs encod  
 A:Reference number: I38849; MUID:95140419; PMID:7838529  
 A:Accession: I38849  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-238 <RES>  
 A:Cross-references: EMBL:U14187; NID:G642832; PIDN:AAC50078.1; PID:G642833  
 C:Gene: CESP:CO3B1.9

A:Gene: GDB:EPLG3  
 A:Cross-references: GDB:438336; OMIM:601381  
 A:Map position: 1q21-1q22  
 C:Superfamily: axon guidance signal protein

Query Match 1.4%; Score 7; DB 2; Length 238;  
 Best Local Similarity 100.0%; Pred. No. 93;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 148 GPGGAE 154  
 |||||  
 Db 78 GPGGAE 84

Search completed: August 27, 2003, 18:45:53  
 Job time : 72 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: September 5, 2003, 08:38:48 / Search time 5500 Seconds  
(without alignments)  
3733.933 Million cell updates/sec

Title: US-10-088-045-2

Perfect score: 2626  
Sequence: 1 MSKPTLITKTLICALSALML.....NKGGQLSADVDVAKAKTKRPN 502

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2888711 seqs, 2045461386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame\_p2n\_model -DEV=x1h  
-O=/cgn2.1/USPRO/spool/US10088045/runat\_04092003\_083142\_3916/app\_query.fasta\_1.647  
-DB=genemb1 -QFMT=fastcap -SUFFIX=rg -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blomsu62 -TRANS=human4.cdi -LIST=45  
-DOCALLIG=200 -THR SCORE=DCT -THR MAX=100 -THR MIN=0 -ALIGN=40 -MODE=LOCAL  
-OUTFMT=pcio -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US10088045 @CGCN 1.1 3508 @runat\_04092003\_083142\_3916 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEBUFFER -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGIO  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmb1:\*  
1: gb\_ba:\*  
2: gb\_hcg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pac:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_scs:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pac:\*  
24: em\_pl:\*  
25: em\_ph:\*  
26: em\_ro:\*  
27: em\_scs:\*  
28: em\_un:\*

29: em\_vl:\*  
30: em\_hcg\_hum:\*  
31: em\_hcg\_inv:\*  
32: em\_hcg\_other:\*  
33: em\_hcg\_mus:\*  
34: em\_hcg\_pln:\*  
35: em\_hcg\_rtd:\*  
36: em\_hcg\_mam:\*  
37: em\_hcg\_vrt:\*  
38: em\_sy:\*  
39: em\_hcgo\_hum:\*  
40: em\_hcgo\_mus:\*  
41: em\_hcgo\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2626	100.0	1506	6	AX098868 Sequence
2	2626	100.0	1509	6	AX098866 Sequence
3	2626	100.0	1759	6	AX394864 Sequence
4	2626	100.0	66986	6	AX067454 Sequence
5	1495.5	53.1	166050	1	AL646085 Ralstonia
6	1394	53.1	97972	1	AE002512 Neisseria
7	1394	53.1	326501	1	AL162757 Neisseria
8	1394	53.1	349980	6	AX044033 Sequence
9	1392	53.0	2069	1	MG97926 Neisseria g
10	787	30.0	3675	1	AB076606 Hypophicr
11	759	28.9	2506	1	AB076606 Halorcul
12	754.5	28.7	1439	1	HD557012
13	494.5	18.8	1823	1	AB046603 Alcalligen
14	486	18.5	2375	1	AB013078 R5U62291
15	474	18.1	1307	1	ACVITRD
16	469.5	17.9	1903	1	AE007256 Slnorhizo
17	456	17.4	10587	1	AE007256 Pseudomonas
18	453	17.3	1886	1	AE007256 Pseudomonas
19	453	17.3	4029	1	AF040987 Agrobacte
20	453	17.3	10484	1	AE008247 Agrobacte
21	453	17.3	10753	1	AE009367 Agrobacte
22	453	17.3	20761	1	AF083948 Pseudomon
23	452	17.2	1709	1	AF051831 Alcalligen
24	451	17.2	1143	1	PAIRKA
25	442	16.8	298550	1	AP005961 Bradyrhiz
26	440	16.8	1499	1	RHU65658 Rhizobium h
27	436.5	16.6	2147	1	BUJ2516
28	431	16.4	11630	1	AE009732 Brucella
29	429	16.3	11302	1	AE014526 Brucella
30	428.5	16.3	316050	1	AE014526 Brucella
31	427.5	16.3	2351	1	AFANIR
32	390.5	14.9	313200	1	AP005214 Corynebac
33	358	9.8	10159	1	AE005046 Halobacte
34	258	9.8	10485	1	AE011334 Leptospir
35	249	9.5	537	1	AF339049 Nitrosomo
36	248.5	9.5	343250	1	AP003594 Nitrosomo
37	245	9.3	347456	1	AE017003 Bacillus
38	232	8.8	537	1	AY078254 Mesorhizo
39	229	8.7	537	1	AY078247 Ensifer s
40	228	8.7	537	1	AY078248 Ensifer s
41	227	8.6	299600	1	AP005942 Bradyrhiz
42	226.5	8.6	540	1	AF339044 Nitrosomo
43	224.5	8.5	540	1	AF339046 Nitrosomo
44	224.5	8.5	540	1	AF339047 Nitrosomo
45	224	8.5	537	1	AY078255 Pseudomon

RESULT 1

## ALIGNMENTS

AX098868  
LOCUS AX098868 1506 bp DNA linear PAT 02-APR-2001  
DEFINITION Sequence 3 from Patent WO0119996.  
ACCESSION AX098868  
VERSION AX098868.1 GI:13538108  
KEYWORDS  
SOURCE Moraxella catarrhalis  
ORGANISM Moraxella catarrhalis  
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
Moraxellaceae; Moraxella.

REFERENCE  
1  
AUTHORS Thomard, J.  
TITLE Moraxella catarrhalis antigen, corresponding gene and uses thereof  
JOURNAL Patent: WO 0119996-A 3 22-MAR-2001;  
SMITHKLINE BEECHAM BIOLOGICALS S.A. (BE)  
FEATURES  
source location/Qualifiers  
1.1506  
/organism="Moraxella catarrhalis"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:480"

BASE COUNT 431 a 356 c 363 g 356 t  
ORIGIN

Alignment Scores:  
Pred. No.: 1.18e-166 Length: 1506  
Score: 2626.00 Matches: 502  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: Gaps: 0

US-10-088-045-2 (1-502) x AX098868 (1-1506)

QY 1 MetSerLySProThrLeuIleValThrThrThrLeuIleCysAlaLeuSerAlaLeuMetLeu 20  
DB 1 ATCTCTAAGCTCTATTTGATTAACCACTTAATTTGCTTAAGTGCATTTGATGCTC 60  
QY 21 SerGlyCySeSerAnGlnAlaAspLysAlaAlaGlnProLysSerSerThrValAspAla 40  
DB 61 AGTGGTTGTGACATCAAGCGACAAAGCCGCCACCAAAAGCAGCAGTGAAGCT 120  
QY 41 AlaAlaLysThrAlaAsnAlaAspAsnAlaAlaSerGlnGlnIleGlnGlyLeuPro 60  
DB 121 GCCGCCAAGACACCAATGAGATTAATGCTGCTTCAAGAACATCAAGCGAGCTGCT 180  
QY 61 ValIleAspAlaIleValThrHisAlaProGluValProProProValAspAspHis 80  
DB 181 GTATTGATGTCATTTGATGACATGACACGAAAGTTCCACACCTGTTGACCGTGAAC 240  
QY 81 ProAlaLysValAlaValLysMetGluThrValGluLysValMetArgLeuAlaAspGly 100  
DB 241 CCCGCCAAGTGTGGTGAATAATGAAACCGTTGAAATAATGCTGCTGCGACATGCG 300  
QY 101 ValGluTyrGlnPheThrPheGlyGlyGlnValProGlyGlnMetIleArgValArg 120  
DB 301 GTGGATATCATGTTTGGACATTTGGCGGTCAAGTTCCAGGCGAGATGATTCGTGCG 360  
QY 121 GluGlyAspThrIleGluValGlnPheSerAsnHisProAspSerLysMetProHis 140  
DB 361 GAAAGGCAACCATCGAAGTGCAGTTCTCAACACCCAGATTCAAAAATGCCCATTAAT 420  
QY 141 ValAspPheHisAlaAlaThrGlyProGlyGlyGlyAlaGluAlaSerPheThrAlaPro 160  
DB 421 GTTGAATTTCAAGCTGCCACAGGCGCTGGCGGCGGCGAGAACGTCATTTACCGAC 480  
QY 161 GlyHisThrSerThrPheSerPheLysAlaLeuGlnProGlyLeuTyrValTyrHisCys 180  
DB 481 GGTTCATACATCAACCTTTAGTTTAAGCCTTACAGCTGTTGATGCTATCACTG 540  
QY 181 AlaValAlaProValGlyMetHisIleAlaAsnGlyMetTyrGlyLeuIleLeuValGlu 200  
DB 541 GGGGTGGCCCTGTTGGCATGACATGCTAATGGCATGATGATTTGATTTGGTTGAA 600

QY 201 ProLysGluGlyLeuProLysValAspLysGluTyrValMetGlnGlyAspPheTyr 220  
DB 601 CCAAAAGAGGGCTTGCCTCAAAAGTAGTAAGATTAATGCTATGACAGGCGACTTTAT 660  
QY 221 ThrLysGlyLysTyrGlyGlnGlnGlyLeuGlnProPheAspMetGluLysAlaIleArg 240  
DB 661 ACCAAAGGCAAAATATGCGCAACAGGCTTACAGCCCTTGTATGTGAAAAGCCATTCGA 720  
QY 241 GluAspAlaGluTyrValPheAsnGlySerValAlaLeuThrGlyGluAsnAla 260  
DB 721 GAAAGTGTGAATATGTTGCTTAAATGATGCGTGGGCGCTTGACTGCTGAAAATGCT 780  
QY 261 LeuLysAlaLysValGlyGluThrValArgLeuPheValGlyAsnGlyGlyProAsnLeu 280  
DB 781 CTAAAGCCAAAGTTGCGCAACCTGCTTCCCTTATTTGTGGTAAACGCGCGCGAATTTG 840  
QY 281 ThrSerSerPheHisValIleGlyGluIlePheAspLysValHisPheGluGlyGlyLys 300  
DB 841 ACATCATCATTCATGCTCATGCTGATGATTTTGTATGAAGTTCACTTGAAGGCTGAAG 900  
QY 301 GlyLysAsnHisAsnIleGlnThrThrLeuIleProAlaGlyValAlaIleThrGlu 320  
DB 901 GGTGAACCAACCAATATCCAAACCAAGCTAATCCAGAGTGGCGCTCCATCACTGAA 960  
QY 321 PheLysValAspValProGlyAspTyrValLeuValAspHisAlaIlePheArgAlaPhe 340  
DB 961 TTTAAGTGTGATGTGCGGCTATTAATGCTTGTGTTGACCAATGCCATCTTCGCTGCTT 1020  
QY 341 AsnLysGlyValAlaLeuGlyLysLeuLysValGluGlyGluAsnHisGluIleTyrSer 360  
DB 1021 AACAAAGGGCATTTGGGCTACTTAAGTGAAGGAGGAAGAAATCATGATGATTTATCA 1080  
QY 361 HisLysGlnThrAspAlaIleTyrLeuProGluGlyValProGlnAlaIleAspThrGln 380  
DB 1081 CACAAACCAACAGACGCTGCTATCTGACAGAGGCGCCCAACCAATGATTAACCCA 1140  
QY 381 GluAlaProLysThrProAlaProAlaAsnGlnGluIleLysAlaGlyLysAla 400  
DB 1141 GAAAGCAACCAAAACCTGCTCACCTGCCAATTAACAAGACAGATTAACAGGTAAAGCA 1200  
QY 401 ThrTyrAspSerAsnGlyAlaIleCysHisGlnProAspGlyLysGlyValProAsnAla 420  
DB 1201 ACCATGATCTTAACGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1260  
QY 421 PheProProLeuAlaAsnSerAspTyrLeuAsnAlaAspHisAlaArgAlaIleSerIle 440  
DB 1261 TTCCACCGCTTGCAACTGTGACTATCTGAACGCCGACACAGCTGCGCCAGCATC 1320  
QY 441 ValAlaAsnGlyLeuSerGlyLysIleThrValAsnGlyAsnGlnTyrGlySerValMet 460  
DB 1321 GTGGCAATGCAATTTGCTGTAAAGATTACCGTCATGCGAACCAATATGAAGCGTCA 1380  
QY 461 ProAlaIleAlaLeuSerAspGlnGlnIleAlaAsnValIleThrTyrThrLeuAsnSer 480  
DB 1381 CTTGGGATTTGCTCTAGACCAACAGATGCGCAATGTCATACCTACACGCTTAACAG 1440  
QY 481 PheGlyAsnLysGlyGlyGlnLeuSerAlaAspAspValAlaLysAlaLysThrLys 500  
DB 1441 TTTGGTAAACAAAGGCGGTCACTCACTGACAGATGTGGCAAAAGCCAAAACCAAG 1500  
QY 501 ProAsn 502  
DB 1501 CCAAAC 1506

RESULT 2  
AX098866 1509 bp DNA linear PAT 02-APR-2001  
LOCUS AX098866  
DEFINITION Sequence 1 from Patent WO0119996.  
ACCESSION AX098866  
VERSION AX098866.1 GI:13538107  
KEYWORDS  
SOURCE Moraxella catarrhalis  
ORGANISM Moraxella catarrhalis

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
FEATURES

1  
Thomard, J.  
Moraxella catarrhalis antigen, corresponding gene and uses thereof  
Patent: WO 011996-A 1 22-MAR-2001;  
SMITHKLINE BEECHAM BIOLOGICALS S.A. (BE)  
Location/Qualifiers

source  
1. 1509  
/organism="Moraxella catarrhalis"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:480"  
BASE COUNT 432 a. 356 c 364 g 357 t  
ORIGIN

Alignment Scores:  
Score: 1,186-166 Length: 1509  
Percent Similarity: 2626.00 Matches: 502  
Best Local Similarity: 100.00% Conservative: 0  
Query Match: 100.00% Mismatches: 0  
Indels: 0  
Gaps: 0

US-10-088-045-2 (1-502) x AX098866 (1-1509)

Qy 1 MetSerLysProThrLeuIleuLysThrThrLeuIleCysAlaLeuSerAlaLeuMetLeu 20  
Db 1 AAGTCAAGACCTACTTGTATTAACACCTTAATTTGTGCTTAAGTCACTTGAAGTGC 60  
Qy 21 SerGlyCysSerAsnGlnAlaAspLysAlaAlaGlnProLysSerSerThrValaAla 40  
Db 61 AGTGGTTGTAGCAATCAACCGGCAAGCGCCACCAAAAAGCAGCGTAGACGCT 120  
Qy 41 AlaAlaLysThrAlaAsnAlaAspAsnAlaAlaSerGlnGlnHisGlnGlyLeuPro 60  
Db 121 GCCGCCAAGACAGCAATGCAATATGCTGCTCAAGAACATCAAGCGCAGCTGCT 180  
Qy 61 ValIleAspAlaIleValThrHisAlaProGluValProProValAspArgAspHis 80  
Db 181 GTCAATTGATGCCATTTGTACGATGACAGAAAGTTCCACACCTTTGACCGTGACAC 240  
Qy 81 ProAlaLysValValValLysMetGluThrValGlnLysValMetArgLeuAlaAspGly 100  
Db 241 CCCGCCAAAGTGGTGTAAATAATGGAACCGTTGAAATAATGTCGCTGCGCAGATGCG 300  
Qy 101 ValGluTyrGlnPheThrPheGlyGlyGlnValProGlyGlnMetIleArgValArg 120  
Db 301 GTGGAATATCAGTTTGGACATTTGGCGGTCAGATTCAGAGGCAATGATTCGTGCGT 360  
Qy 121 GluGlyAspThrIleGluValGlnPheSerAsnHisProAspSerLysMetProHisAsn 140  
Db 361 GAAGCGACACCATGCAAGTGCAGTTCTCAACACCAAGATTCAAAAATGCCCATAT 420  
Qy 141 ValAspPheHisAlaAlaThrGlyProGlyGlyGlyAlaGlnLysSerPheThrAlaPro 160  
Db 421 GTTGACTTTCAAGCTCCACAGGCGCTGGGCGGCGGAGAGCGTCATTATACCGACCG 480  
Qy 161 GlyHisThrSerThrPheSerPheLysAlaLeuGlnProGlyLeuTyrValTyrHisCys 180  
Db 481 GGTCAATACATCAACCTTTAGTTTAAAGCCCTTACACCTGTTGATGCTATCACTGT 540  
Qy 181 AlaValAlaProValGlyMetHisIleAlaAsnGlyMetTyrGlyLeuIleLeuValGlu 200  
Db 541 GCGGTTGCCCTGTTGGCATGCACTTGTCAATGAGCATGATAGTTGATTTGGTTGA 600  
Qy 201 ProLysGluGlyLeuProLysValAspLysGlyTyrTyrValMetGlnGlyAspPheTyr 220  
Db 601 CCAAAAAGAGGCTTGCCTTAAAGATGATTAAGATCTATGTCAGACGCGACTTTAT 660  
Qy 221 ThrLysGlyLysTyrGlyGlnGlnGlyLeuGlnProPheAspMetGluLysAlaIleArg 240  
Db 661 ACCAAAGGCAAAATATGCGCAAGGCTTACAGCCCTTGAATATGGAATAAAGCCATTGCA 720

Qy 241 GluAspAlaGluTyrValValPheAsnGlySerValGlyAlaLeuThrGlyGlyLysAla 260  
Db 721 GAAGATGCTGAATATGTTGCTTTAATGTTTCGGGGGCGCTTCACTGTCGCAAAATGCT 780  
Qy 261 LeuLysAlaLysValGlyGluThrValArgLeuPheValGlyAsnGlyGlyProAsnLeu 280  
Db 781 CTAAAGCCCAAGTTGGGCAACCTGTTATTTGGGTTAAGGCGGCGCCCAATTTTG 840  
Qy 281 ThrSerSerPheHisValIleGlyGluIlePheAspLysValHisPheGluGlyLys 300  
Db 841 ACATCATCTATTCATGCTATGTTGATGAGATTTTGTATTAAGTTCACTTTGAGGCTGAAG 900  
Qy 301 GlyLysAsnHisAsnIleGlnThrThrLeuIleProAlaGlyGlyAlaAlaIleThrGlu 320  
Db 901 GGTGAAAACCAATATCCAAACCAAGCTTAATCCAGAGGTGGCGCTGCCATCTGAA 960  
Qy 321 PheLysValAspValProGlyAspTyrValLeuValAspHisAlaIlePheArgAlaPhe 340  
Db 961 TTTAAGGTGATGTCGGGTGATATGCTTGGTTGACCATGCACTTCCGTCCTTT 1020  
Qy 341 AsnLysGlyAlaLeuGlyIleLeuLysValGlyGlyGluLysAsnHisGluIleTyrSer 360  
Db 1021 AACCAAGGGCATTTGGCATCTTAAGTGAAGTGAAGGAAATCATGAGATTTATTCA 1080  
Qy 361 HisLysGlnThrAspAlaValTyrLeuProGluGlyAlaProGlnAlaIleAspThrGln 380  
Db 1081 CACAAACAAACAGACGCTGTATCTGCGCAGAGGTGCCCAAGCAATGATACCCAA 1140  
Qy 381 GluAlaProLysThrProAlaProAlaLeuGlnGlnGlnIleLysAlaGlyLysAla 400  
Db 1141 GAAGACCCCAAAACCTGACCTGCACTTCAACAGAGCAATTAAGACAGTAAAGCA 1200  
Qy 401 ThrTyrAspSerAsnCysAlaAlaCysHisGlnProAspGlyLysGlyValProAsnAla 420  
Db 1201 ACTTATGACTTCACTGCTGCTGCTGCTGCTGCAACCACTGATGTAAGGCTGCCAAACGT 1260  
Qy 421 PheProProLeuAlaAsnSerAspTyrLeuAsnAlaAspHisAlaArgAlaAspSerIle 440  
Db 1261 TTCACACGCGCTGCCAATCTGATCTGAACGCGGCAAGCGCTGTCGCCCGACATC 1320  
Qy 441 ValAlaAsnGlyLysSerGlyLysIleThrValAsnGlyAsnGlyTyrGlnSerValMet 460  
Db 1321 GTGGCAATATGATTTCTGTTAAGATTACCGGCAATGCGCAATATGAAGCGTCAG 1380  
Qy 461 ProAlaIleAlaLeuSerAspGlnGlnIleAlaAsnValIleThrTyrThrLeuAsnSer 480  
Db 1381 CTTGCGATTTGCTTACAGGACCAACAGATTCATCATCACTTACACGCTTAACAGC 1440  
Qy 481 PheGlyAsnLysGlyGlyLysLeuSerAlaAspAspValAlaLysAlaLysLysThrLys 500  
Db 1441 TTTGGTAAACAAAGCGCTCAACTCAGTGCAGACGATGTGGCAAAAGCCAAAACCAAG 1500  
Qy 501 ProAsn 502  
Db 1501 CCAAAC 1506  
RESULT 3  
AX394864 1759 bp DNA linear PAT 18-MAY-2002  
LOCUS AX394864  
DEFINITION Sequence 1 from Patent WO0218595.  
ACCESSION AX394864  
VERSION AX394864.1 GI:21065937  
KEYWORDS  
SOURCE  
ORGANISM  
Moraxella catarrhalis  
Moraxella catarrhalis  
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
Moraxellaceae; Moraxella.  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
1  
Loosmore, S., Wang, J., Bradley, B., Ochs, M., and Yang, Y. P.  
Moraxella polypeptides and corresponding dna fragments and uses  
thereof  
Patent: WO 0218595-A 1 07-MAR-2002;  
Aventis Pasteur Limited (CA)

## FEATURES

source

## CDS

Location/Qualifiers  
 1..1759  
 /organism="Moraxella catarrhalis"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:480"  
 101..1609  
 /note="unnamed protein product"  
 /codon\_start=1  
 /transl\_table=1  
 /protein\_id="CAD32058.1"  
 /db\_xref="GI:21065938"

/translation="MSKPTLTKITLTCALSLMLSGSNQADKAQPKSTVDAAKT  
 ANDNASQEHQBLEPVIDAVTHAPVPPVDHDPKAVVMEVEKWRADVE  
 YQFTGQVQGMIRREGDTIEVQSNPDSKPNVDPAHATGPGGGAESFTAP  
 GHTSTEFKALQGLVYHCAVAPVGMHINGMVGLIVPEKGLPKVDKEYVMGD  
 FYTKKXGEQGLQPFDMKEKAREDAEYVINGSVGLATGENALKAKGVTRELVNG  
 GPULTSFHYIGELPDKHFEGGKGENHIOITLLIPAGSAITREPVVDGDIYLDH  
 AIRPAPKKAAGILKVEENHETYSKQTDANYLPBGAQALIDTQAPTPPAAID  
 EQIRAGATYDSNCAACHQPDGKGVPAFPPLANSIDYLNADHARLAASYANGISGIT  
 VNGQYESVMPALALSDQILANVITYTLNLSFGNKGQQLSADVAKAKTKPN"

BASE COUNT 497 a 396 c 414 g 452 t  
 ORIGIN

## Alignment Scores:

Pred. No.: 1,41e-166 Length: 1759  
 Score: 2626.00 Matches: 502  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 6 Gaps: 0

US-10-088-045-2 (1-502) x AX394864 (1-1759)

QY 1 MetSerLyProThrLeuIleLyThrThreuleCySaIaLeuSerAlaLeuMetLeu 20  
 DB 101 ATGCTTAAGCTTACTTTGATTAACCACTTAATTTTGCTTAAGTGCATTGATGCTC 160  
 QY 21 SerGlyCySerSerArginalaAspLySaIaIaGlnProLySerSerThrValaAspAla 40  
 DB 161 AGTGGTGTGATCATCAAGCGGCAAGCGCCGACCAAAAAGCAGCAGTGAAGCGCT 220  
 QY 41 AlaAlaLyThrAlaAsnAlaAspAsnAlaIaSerGlnGlnIleGlnIleGluLeuPro 60  
 DB 221 GCGCCCAAGACACCAATGAGATTAATGCTCTCAACAAACATCAAGCGGCGCTGCT 280  
 QY 61 ValIleAspAlaIleValThrAlaIaProGluValaProProProValaAspAspAla 80  
 DB 281 GTCAATGATGCCATTGTTACGATGCACCAAGAGTTCCACCACTGTTGACCGTACAC 340  
 QY 81 ProAlaLyValaValaLySerGluThrValaGluLyValaMetArgLeuAlaAspGly 100  
 DB 341 CCGCCCAAGGTGGGTAAATGAGAAACCGTTGAAAAGTCAATGCGTCTGGCAGATGGC 400  
 QY 101 ValGluLyGlnPheThrPhePheGlyGlnValaProGlyGlnMetIleArgValaArg 120  
 DB 401 GTGGAATATAGTTTGACATTTGGCGGTCAAGTTCCAGGCGAGATGATTCGTGTCGT 460  
 QY 121 GluGlyAspThrIleGluValaGlnPheSerAsnHisProAspSerLySerMetProHisAsn 140  
 DB 461 GAAGGCGACCAATCGAAGTGCAGTTCTCAAAACCAACCAATTCAAAATGCCCAATAT 520  
 QY 141 ValAspPheHisAlaAlaThrGlyProGlyGlyValaGluValaSerPheThrAlaPro 160  
 DB 521 GTTGACTTTCACGCTGCACAGGGCGTGGCGGGGCAAGACGTCATTTACCGACCG 580  
 QY 161 GlyHisThrSerThrPheSerPheLyAlaLeuGlnProGlyLeuTyValIleGlyArg 180  
 DB 581 GGTCAATACATCAACTTTAGTTTAAAGCCTTACAGCTGTTGATGCTATCACTG 640  
 QY 181 AlaValaIaProValaGlyMetHisIleAlaAsnGlyMetTyGlyLeuIleLeuValaGlu 200  
 DB 641 GCGGTGGCCCTGTGGCATGACATTTGTAAGGATGATGATGATTTGTTGTTGA 700

QY 201 ProLyGluGlyLeuProLyValaAspLyGluTyTyValaMetGlnIleAspPheTy 220  
 DB 701 CCAAAAAGAGCGCTTGCCAAAGTAGTAAGAAATAGTATGATGATGATGATGATGATGAT 760  
 QY 221 ThrLyGlyLyTyGlyGlnIleGlnIleGlnIleGlnIleGlnIleGlnIleGlnIle 240  
 DB 761 ACCAAAGGCAATATGCGCAACAGGCTACAGCCCTTGAATGAGAAAAGCCATTGCA 820  
 QY 241 GluAspAlaGluTyValaValaPheAsnGlySerValaGlyAlaLeuThrGlyGluAsnAla 260  
 DB 821 GAAGATGCTGAATATGTTGCTTTAAATGTTGCTGGGGGCGTTGATGCTGTAATAATGCT 880  
 QY 261 LeuLyAlaLyValaGlyGlnIleGlnIleGlnIleGlnIleGlnIleGlnIleGlnIle 280  
 DB 881 CTAAAAAGCCAAAGTTGGCCAAACCTGCTTATTTGATGAGTAAAGCGCGCCCAATTGG 940  
 QY 281 ThrSerSerPheHisValaIleGlyGlnIlePheAspLyValaHisPheGlnIleGlyArg 300  
 DB 941 ACATCATCTTCATCATGCTCATGCTGATGATTTTGAATGAGTTTCACTTGAGGGTGTAAAG 1000  
 QY 301 GlyGluAsnHisAsnIleGlnThrThreuleIleProAlaGlyAlaIleAlaIleThrGlu 320  
 DB 1001 GGTGAAAACCAACATATCCAAACCAACGCTAATCCACAGAGTGGCGCTGCCATCACTGA 1060  
 QY 321 PheLyValaAspValaProGlyAspTyValaLeuValaAspHisAlaIlePheAlaPhe 340  
 DB 1061 TTTAAGTGTGATGCTCGCGGTGATTAATGCTGTTGACCATGCTTCCTGCTCTTT 1120  
 QY 341 AsnLyGlyValaLeuGlyTyIleLeuLyValaGlyGlyGluAsnHisGlnIleTySer 360  
 DB 1121 AACAAAGGGGCAATGGGCACTTAAAGTGAAGGTAAGAAATCATGAGATTAATCA 1180  
 QY 361 HisLyGlnThrAspAlaValaTyLeuProGlyAlaProGlnAlaIleAspThrGln 380  
 DB 1181 CACAAACCAACAGAGCTGTCTATCTGCCAGAGGTTGCCCAACCAATTAATCA 1240  
 QY 381 GluAlaProLyThrProAlaProAlaLeuGlnGlnIleLyAlaGlyLyAla 400  
 DB 1241 GAAGCAACCAACACCTGCACTGCACTTCAACAGACAGATTAAGCAGATTAAGCA 1300  
 QY 401 ThrTyAspSerAsnCyAlaAlaCyHisGlnProAspGlyLyGlyValaProAsnAla 420  
 DB 1301 ACCTATGATCTTAACGCTGCTGCTTGTCAACCACTGATGTAAGCGCTGCAACCGCT 1360  
 QY 421 PheProProLeuAlaAsnSerAspTyTyLeuAsnAlaAspHisAlaArgAlaIleSerIle 440  
 DB 1361 TTCACACCGCTTGCCAACTGATCTGATCTGAAACGCCAGCAGCTCGTCCGCGCAGATC 1420  
 QY 441 ValaIaAsnGlyLeuSerGlyLyIleThrValaAsnGlyAsnGlnTyTyIleSerValaMet 460  
 DB 1421 GTGGCAAAATGATGTTGCTGTGAATTAACCGTCATGCAATGCAATATGAAGCGTCAATG 1480  
 QY 461 ProAlaIleAlaLeuSerAspGlnIleIleAlaAsnValaIleThrTyThrLeuAsnSer 480  
 DB 1481 CCTGGATTTGCTTACGACCAACAGATGTCATGATCACTTACCGTTAAACGC 1540  
 QY 481 PheGlyAsnLyGlyGlyGlnLeuSerAlaAspAspValaAlaLyAlaLyTySerThyLy 500  
 DB 1541 TTTGATACCAAGGCGGTCACTCACTGACGACGATGTGGCAAAAGCCAAAACCAAG 1600  
 QY 501 ProAsn 502  
 DB 1601 CCAAC 1606

RESULT 4  
 AX067454  
 LOCUS AX067454 66986 bp DNA linear PART 24-JAN-2001  
 DEFINITION Sequence 29 from Patent WO0078968.  
 ACCESSION AX067454  
 VERSION AX067454.1 GI:12545074  
 KEYWORDS  
 SOURCE Moraxella catarrhalis  
 ORGANISM Moraxella catarrhalis

Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
Moraxellaceae; Moraxella.  
1  
REFERENCE  
AUTHORS  
TITLE  
Nucleotide sequences of moraxella catarrhalis genome  
JOURNAL  
Patent: WO 0078968-A 29 28-DEC-2000;  
Incyte Genomics, Inc. (US)  
FEATURES  
Location/Qualifiers  
source  
1. 66986  
/organism="Moraxella catarrhalis"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:480"

BASE COUNT 18889 a 13427 c 15112 g 19558 t  
ORIGIN

## Alignment Scores:

Pred. No.: 8, 88e-165 Length: 66986  
Score: 2626.00 Matches: 502  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
Gaps: 0

US-10-088-045-2 (1-502) x AX067454 (1-66986)

QY 1 MetSerLysProThrLeuIleLysThrThrLeuIleCysAlaLeuSerAlaLeuMetLeu 20  
DB 3742 ATGTCTAAGCTTACTTGTATTAACAACCTTAATTGTGCTTAAGTCACTTGAATGCC 3801  
QY 21 SerGlyCysSerAsnGlnAlaAspLysAlaAlaGlnProLysSerSerThrValAspAla 40  
DB 3802 AGTGGTTGTAGCAATCAACGCGAACAAGCCGCCCAAAAAGCAGCGGTAGACGCT 3861  
QY 41 AlaAlaLysThrAlaAspAlaAspAlaAlaSerGlnGluHisGlnGlyLysLeuPro 60  
DB 3862 GCCGCGAAGACAGCAATGCAATATGCTGCTCAACAAGATCAATCAAGCGAGCTGCT 3921  
QY 61 ValIleAspAlaIleValThrHisAlaProGluValProProValAspArgAspHis 80  
DB 3922 GTCAATGATGCATTTGATGCGATGACCAAGATTCACCACTGTTGACCGTGACAC 3981  
QY 81 ProAlaLysValAlaValLysMetGlnThrValGluLysValMetArgLeuAlaAspGly 100  
DB 3982 CCCGCCAAAGTGGTGTAAATAATGAAACCGTTGAAATAAGTCATGGCTGCGCAATGCG 4041  
QY 101 ValGluThrGlnPheThrPheGlyGlyGlnValProGlyGlnMetIleArgValArg 120  
DB 4042 GTGGAATATCAGATTTTGACATTTGGCGGTCAAGTTCCAGGCGAGATGATGCTGCGCT 4101  
QY 121 GlnGlyAspThrIleGluValGlnPheSerAsnHisProAspSerLysMetProHisAsn 140  
DB 4102 GAAGCGACACCATCGAAGTGCATTTCTAAACCCAGATTCAAAATATGCCCATTAAT 4161  
QY 141 ValAspPheHisAlaAlaThrGlyProGlyGlyAlaGluAlaSerPheThrAlaPro 160  
DB 4162 GTTGACTTTCACGCTGCCACAGGCGCTGGCGGGCGAAGCGTCATTTACCGGACCG 4221  
QY 161 GlyHisThrSerThrPheSerPheLysAlaLeuGlnProGlyLysLeuThrValIleArg 180  
DB 4222 GGTCTATCACTCAACCTTTAGTTTAAAGCCTTAACGCTGTTGATCTCATCTCTGT 4281  
QY 181 AlaValAlaProValAlaGlyMetHisIleAlaAsnGlyMetThrGlyLeuIleLeuValGlu 200  
DB 4282 GCGGTGGCCCTGTTGGCATGCACTGCTTAATGCGATGATGTTGATTTGGTTGA 4341  
QY 201 ProLysGluGlyLysProLysValAspLysGluThrTyrValMetGlnIleAspPheThr 220  
DB 4342 CCAAAAAGAGGCGCTGCGCAAAAGTAAAGATATGATGTCAGCAAGCGCATTTTAT 4401  
QY 221 ThrLysGlyLysThrGlyGlnGlnGlyLeuGlnProPheAspMetGlnLysAlaIleArg 240  
DB 4402 ACCAAAGGCAAAATATGCGCAACAAGGTCTACAGCCCTTTGATATGAAAAAACCATTTCA 4461

QY 241 GlnAspAlaGluThrValAlaPheAsnGlySerValGlyAlaLeuThrGlyGluAsnAla 260  
DB 4462 GAAGATGCTGAATATGTTGCTTAAATGCTTCGGTGGCGGCTGATCGTGAATAATGCT 4521  
QY 261 LeuLysAlaLysValGlyGlnThrValArgLeuPheValGlyAsnGlyGlyProAsnLeu 280  
DB 4522 CTAAAAGCGAAGTTGGGAAACGTGCTTATTTGGGTAAAGCGGCGCCGCAATTTG 4581  
QY 281 ThrSerSerPheHisValIleGlyGluIlePheAspLysValHisPheGlyGlyLys 300  
DB 4582 ACATATATATTCATGATGATGTCAGATTTTGTATTAAGTTCACTTTGAGGCTGTAAG 4641  
QY 301 GlyGluAsnHisAsnIleGlnThrThrLeuIleProAlaGlyGlyAlaAlaIleThrGlu 320  
DB 4642 GGTGAATAACCAATATCCAAACCAAGCTTAATCCAGAGGCGGCGCTGCATCACTGA 4701  
QY 321 PheLysValAspValProGlyAspThrValLeuValAspHisAlaIlePheArgValPhe 340  
DB 4702 TTTAAGGTGATGTCGCCGGGTGATTAATGCTTTGGTTCACATGCTTCGTCCTTT 4761  
QY 341 AsnLysGlyAlaLeuGlyIleLeuLysValGluGlyGluGluAsnHisGluIleTyrSer 360  
DB 4762 AACCAAGGCGCATTTGGCATCTTAAGTGCAGAGGTGAAGAAATCATGATTTATTTCA 4821  
QY 361 HisLysGlnThrAspAlaValTyrLeuProGluGlyAlaProGlnAlaIleAspThrGln 380  
DB 4822 CACAACCAACAGACGCTGCTATCTGCGCAGAGGCGCCCAACAAGCATTTGATCCCA 4881  
QY 381 GluAlaProLysThrProAlaProAlaAsnLeuGlnGlnIleLysAlaGlyLysAla 400  
DB 4882 GAAGCACCACCAACCTGACCTGACCTTAACGACGAGATTAAGCAGTTAAGCA 4941  
QY 401 ThrTyrAspSerAsnCysAlaAlaCysHisGlnProAspGlyLysGlyValProAsnAla 420  
DB 4942 ACCTATGATCTTAATGCTGCTGCTTTGTACCAACTGATGTAAGGCTGCCAACCT 5001  
QY 421 PheProProLeuAlaAsnSerAspTyrLeuAsnAlaAspHisAlaArgAlaAlaSerIle 440  
DB 5002 TTCCACCGCTTGCCCACTGATGATCTGAACGCGACCAAGCTGCGCCGCGCATTC 5061  
QY 441 ValAlaAsnGlyLysSerGlyLysIleThrValAsnGlyAsnGlnTyrGluSerValMet 460  
DB 5062 GTGGCAATGATGCTGCTGATGATTAACCGTCATATGCAACATATGAAGAAACCTCATG 5121  
QY 461 ProAlaIleAlaLeuSerAspGlnGlnIleAlaAsnValIleThrTyrThrLeuAsnSer 480  
DB 5122 CTGCGATTTGCTGAGGACCAACAGATGCGAATGATCATCACTTAACGCTTAACGCC 5181  
QY 481 PheGlyAsnLysGlyGlyGlnLeuSerAlaAspAspValAlaLysAlaLysLysThrLys 500  
DB 5182 TTTGCTAACAAGGCGTCACACTGATGTCAGATGTCAGATGTCGCAAAAAGCCAAAACCAAG 5241  
QY 501 ProAsn 502  
DB 5242 CCAAC 5247

RESULT 5  
LOCUS AL646085/c 166050 bp DNA linear BCT 11-APR-2003  
DEFINITION Ralstonia solanacearum GMI1000 megaplasmid, complete sequence;  
segment 10/11.  
ACCESSION AL646085 AL646053  
VERSION AL646085.1 GI:17431932  
KEYWORDS  
SOURCE Ralstonia solanacearum  
ORGANISM Ralstonia solanacearum  
Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
Burkholderiaceae; Ralstonia.  
REFERENCE  
1  
AUTHORS  
Salanoubat, M., Genin, S., Artiguenave, F., Gouzy, J., Mangenot, S.,  
Arlat, M., Billault, A., Broctier, P., Camus, D. C., Cattelino, L.,  
Chandler, M., Choisme, N., Claudel-Renaud, C., Cunac, S., Demange, N.,  
Gaspin, C., Lavie, M., Moisan, A., Robert, C., Saurin, W., Schlex, T.,

```

TITLE       Siguiet, P., Thebaud, P., Whalen, M., Wincker, P., Levy, M.,
JOURNAL     Weissenbach, J. and Boucher, C.A.
MEDLINE     Genome sequence of the plant pathogen Ralstonia solanacearum
PUBMED      Nature 415 (6871), 497-502 (2002)
REFERENCE   21681879
AUTHORS      11823852
TITLE        2 (bases 1 to 166050)
JOURNAL      Boucher, C.A.
             Direct Submission
             Submitted (05-DEC-2001) Genoscope and CNRS UMR-8030, 2 rue Gaston
             Cremieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biologie
             Moléculaire des Interactions Plantes-Microorganismes INRA-CNRS,
             BP27, 31326 Castanet-Tolosan Cedex, France, Fondation Jean
             Dausset-CEPH, 27 rue Juliette Dodu, 75010 Paris, France, IMGM CNRS
             118 Route de Narbonne, F 31062 Toulouse Cedex 4, Genoscope and INRA
             URGV, 2 rue Gaston Cremieux, CP5706, 91057 Evry Cedex, France,
             Laboratoire de Biometrie et Intelligence Artificielle INRA, BP27,
             F31326 Castanet-Tolosan Cedex, Laboratoire de Genetique Cellulaire
             INRA, BP27, F31326 Castanet-Tolosan Cedex
             Christian.Boucher@toulouse.inra.fr
             http://sequence.toulouse.inra.fr/R.solanacearum.html.

FEATURES
source      Location/Qualifiers
            1..166050
            /organism="Ralstonia solanacearum"
            /mol_type="genomic DNA"
            /strain="GM11000"
            /db_xref="taxon:305"
            /plasmid="megaplasmid"
            /complement(105..1829)
            /gene="RSP1459"
            /note="synonym: RS03085"
            /complement(105..1829)
            /gene="RSP1459"
            /function="miscellaneous: hypothetical/global homology"
            /note="Product confidence : probable
            Gene name confidence : hypothetical
            predicted by Codon usage
            predicted by FrameD"
            /codon_start=1
            /evidence=not experimental
            /transl_table=11
            /product="PROBABLE SIGNAL PEPTIDE PROTEIN"
            /protein_id="CAD18610.1"
            /db_xref="GI:17431933"
            /db_xref="GOA:O8X027"
            /db_xref="SPTREMBL:O8X027"
            /translation="MSANTGRITGALLLGFAGSPGASAPAEIYVNRGRPTTIDA
            GFWQAVAVRDRFVAISDAEVAQLVGPATQVNDLHGAMVVPGLSDVHAHPLEGAY
            EALYHCALAPTDGLDVLVAVGAARAPDPMIVGPMSSALLGLLETPESLALDR
            ASKGLPVLRLDPTFHNRMANSVYLQAGAGPAGAAAGAYVADGKPTGLIKEPPAW
            GVEKELIPRTPARLGHAAEAARLNSFGITGVODAFTROIITGMHAADTGGCLP
            LNVVSGVPEAGGDCGLPDROTLAVRSRDLRPVVKYFLDGVPMSTYAAMLPAP
            SPANHGDFRGAFLYALAVUTKLTALDRGCTPKLHAVDGAVRLALDLAEVRRANG
            KMGPRHQIHLITFIAPSDMRFRRLNVVADVSPMLFPHAYTPLFEKVVGHARTLSHY
            PIGSLIRAGTLVAGSDPDAGQDTPDWGIBGLVTRNPAGNVPGVLDHAIRLQO
            ALRLYLNSATAMGLQGETGSIEGKSDAFVLSQDLFXIIPSRIHRTVTVTYFGCR
            PYVRADO"
            /complement(2011..2835)
            /gene="RSP1460"
            /note="synonym: RS03084"
            /complement(2011..2835)
            /gene="RSP1460"
            /function="miscellaneous: unknown"
            /note="Product confidence : hypothetical
            Gene name confidence : hypothetical
            predicted by Codon usage
            predicted by FrameD"
            /codon_start=1
            /evidence=not experimental
            /transl_table=11
            /product="HYPOTHETICAL PROTEIN"
            /protein_id="CAD18611.2"
            /db_xref="GI:29823183"

gene
            /db_xref="GOA:O8X026"
            /db_xref="SPTREMBL:O8X026"
            /translation="MKLARTPPSPSSAAGASAEPSAPSPVSGKSPSPGALAYLSQ
            AAMPKRLRLASDGLSAGPEVKA PALPTLIGTRADIDVRPSLQPVPRKDDGI
            HDMEKDAEGRPLAHPRTALOGGDAPPMTDPAVAKADIDALKAGDKQYIWAASLG
            RVFIEEPPVGHDPDSGKRRGHPPLVSGGPARICGSEFPDAETGKLVINKLGRYS
            RYEDRRAQLQEVATIIRAVAAPLGLLEVETEVSGKTDPALPLDPRYRPPAD"
            /complement(3462..3779)
            /gene="RSP1461"
            /note="synonym: RS03083"
            /complement(3462..3779)
            /gene="RSP1461"
            /function="miscellaneous: hypothetical/global homology"
            /note="Product confidence : probable
            Gene name confidence : hypothetical
            predicted by Codon usage
            predicted by FrameD"
            /codon_start=1
            /evidence=not experimental
            /transl_table=11
            /product="PROBABLE SIGNAL PEPTIDE PROTEIN"
            /protein_id="CAD18612.1"
            /db_xref="GI:17431935"
            /db_xref="GOA:O8X025"
            /db_xref="SPTREMBL:O8X025"
            /translation="MHSFGSNRGRKSIYVAALVMASAFAPHAHSAFTLGKVDYGTG
            TALTPANMSATWGNAMDACRVQYCTTRSVMNTHQFGNPAGSNKSVAAWTWCBD
            TAN"
            /complement(3989..5467)
            /gene="RSP1462"
            /note="synonym: RS03082"
            /complement(3989..5467)
            /gene="RSP1462"
            /function="miscellaneous: hypothetical/partial homology"
            /note="Product confidence : hypothetical
            Gene name confidence : hypothetical
            predicted by Codon usage
            predicted by FrameD"
            /codon_start=1
            /evidence=not experimental
            /transl_table=11
            /product="HYPOTHETICAL PROTEIN"
            /protein_id="CAD18613.1"
            /db_xref="GI:17431936"
            /db_xref="GOA:O8X024"
            /db_xref="SPTREMBL:O8X024"
            /translation="WLAAPPFAIAEALPAHPPTARLRLQELQCALTRAVSVQCOQ
            DLQOQEGTEQARRHEFLPQTVLSQLEKNSITSGTGTTRGASSGNLVTWKLNSGT
            QLSSEGLVNRFGSAPADSPVPGASINTSRVTSLSQPLRSGSDVVLNERR
            AALAESARSLFQTRDVLVNGVLAYPEALRQNVTLAQLRAGEARSINLAL
            AAGRARIALLOSDDLQAEALRQSOAEVVARROLRVGRYDLADRDTITTD
            SPATLPPAANEARIPMDALARPDLQADLADAYVAOSFVTAADGMNOLDIYARL
            DRSGDSAGTSTSRQINRNVGVSFSPDLKESRSLALGARAVALTAELSLADLETAR
            AVEUSDALNIFALFQHMMAQRTALARLAGEIEKAPAGSSATDLSLADTINQA
            LSOEQAFPAITFQLEQVSGTALBQMHVAAPQATLPBS"
            /complement(5582..6835)
            /gene="RSP1463"
            /note="synonym: RS03081"
            /complement(5582..6835)
            /gene="RSP1463"
            /function="cell processes; transport of small molecules"
            /note="Product confidence : putative
            Gene name confidence : hypothetical
            predicted by Codon usage
            predicted by FrameD"
            /codon_start=1
            /evidence=not experimental
            /transl_table=11
            /product="PUTATIVE ATP-BINDING ABC TRANSPORTER PROTEIN"
            /protein_id="CAD18614.1"
            /db_xref="GI:17431937"
            /db_xref="GOA:O8X023"
            /db_xref="SPTREMBL:O8X023"

```

```

/translation="MARPAVQERGHMLKARQASSOPILLQSTLAPVDANVIAAEVGE
VLLAKFVQFDITVAGQKLAQVSDTELKQLEAEIASIQAOQALSAQOISSTDEYQ
AARBLIAQNSLIAAQRKEAQOMLYDKGIARTLELDKQOEDITQSOVIGADEI
ATLRKAPALRVLRLLELLEQLRLELRLKATLLSPISGVLYVEPDGNDPS
GKEELNAGAVLPKDVILITVPTSAFLIKAWDEDEIRRVAGQAVRIYEGADSRAR
ACTVORVSSQARAADRGSGSRTAEFEHQILFKPPTDGRSLRGSANMSLAIEPG
PTTLIPLIAVAMSAPGRPVAVRKVSASDTARMQETLEGRITLVDSVEVAGIADGDEV
WPSAAGGPDGTLKQLLSGVNG"
gene
6828..7610
/gene="RSPI464"
/feature="synonym: RS03080"
6828..7610
/gene="RSPI464"
/function="cell processes; transport of small molecules"
/feature="Product confidence: probable"
Gene name confidence: hypothetical
Predicted by Codon_usage
Predicted by Homology
Predicted by "Framed"
/codon_start=1
/evidence=not experimental
/translation=1
/product="PROBABLE ATP-BINDING ABC TRANSPORTER PROTEIN"
/protein_id="CAD18615.1"
/db_xref="GI:17431938"
/db_xref="GOA:Q8XQ22"
/db_xref="SPTREMBL:Q8XQ22"
/translation="MAEPKAGNDRA PAGLLIELROVSRFPVAGETVA RGVLD:OVR
SGEVLAMVCPSSGSKSTMNLLGLDPTPSGOYRLAGVDTQTLSSQESRLRNDQIG
VFOQFHLPLATLQNAQPLPLRGISQSLADAPALVAALVALEHRSHRAELSGG
ERQVAPRAAMAGOPRLILADEPTGLDSATGARVALALREYATHTGALLMTIDLE
IARQVPROIEMDGLRCATQDPRAPAAAPQPLPGRGTVA"
7612..8874
/gene="RSPI465"
/feature="synonym: RS03079"
7612..8874
/gene="RSPI465"
/function="miscellaneous; hypothetical/global homology"
/feature="Product confidence: probable"
Alignment Scores:
Pred. No.: 8.4e-89 Length: 166050
Score: 1495.50 Matches: 294
Percent Similarity: 69.92% Conservative: 71
Best Local Similarity: 56.32% Mismatches: 118
Query Match: 56.95% Indels: 39
DB: 1 Gaps: 9
US-10-088-045-2 (1-502) x AL646085 (1-166050)
QY 12 ILeCYaLaLeuSerAlaLeuMetLeuSerGlyCySerAsn----- 25
DB 50007 GTTGGCTGTGTCTCAAGCTTCAGCTGTGCGCGCTCTTACGTGAGACAGACACAC 49948
QY 26 -----GlnAlaAspLysAlaIaGlnProLysSerSerThrValAspAlaIaIaLys 43
DB 49947 ACCGACGACGCGGAGAAA-----AAATAGAGCTGTCTTCAACGCGTCCGCGT 49900
QY 44 ThrAlaAsnAla-----AspAsnAlaIaSer 52
DB 49899 CTGATGCGCGCGCGCTGATTTGCGTGAAGCTGTGTACTGCTGACGAGCGCGCGGCGC 49840
QY 53 GlnGluHisGlnGlyLeuLeuPro-----ValIleAsp 63
DB 49839 CAG-----GCAGCCAAAGCTTCCCGGCGATTCGCTCCGCGCGCGCGAGCCATCCGC 49786
QY 64 AlaIleValThrHisAlaProGluValProProProValAspArgAspHisProAlaLys 83
DB 49785 GCGGAGCTCAGCAGTCCGCCCAAGTGGCGCGCGCGATCCACGCAATATCCGCGCAG 49726
QY 84 ValIleValLysMetGluThrValGluLysValMetArgLeuAlaAspGlyValGluTyr 103
DB 49725 GTTCATCTGAGACTGGAAAGTATCGAAGAAATGCGATCTCCAGAGGGGTCAAGTAC 49666

```

```

QY 104 GlnPheTrpThrPheGlyGlyGlnValProGlyGlnMetIleArgValArgGlyAsp 123
DB 49665 ACCTCTGGACCTTCGCGCGGACGAGTGCAGGACACTTCATCCGCTGGCGGAGGCGAT 49606
QY 124 ThrIleGluValGlnPheSerAsnHisProAspSerLysMetProHisAsnValPhe 143
DB 49605 ACCGGAGGATTCACCTCAAGAACACCGCGACACCAAGATCCGACAACTCAACCTCG 49546
QY 144 HisAlaAlaThrGlyProGlyGlyGlnValIaGluAlaSerPheThrAlaProGlyHisThr 163
DB 49545 CAGCGCGTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 49486
QY 164 SerThrPheSerPheLysAlaLeuGlnProGlyLeuTyrValTyrHisCySerAlaValAla 183
DB 49485 TCGAGTTCACCTTCAGAGCCCTCAATAGAGGGGCTATGTCTTACACTGCGCGCACCGG 49426
QY 184 ProValGlyMetHisIleAlaAsnGlyMetTyrGlyLeuIleLeuValGluProLysGlu 203
DB 49425 CCGGTGCGCATGCAATATGCGCAACGCGCATGTATGCGCTGATCTGTCGACACCGCGGAA 49366
QY 204 GlyLeuProLysValAspLysGluTyrTyrValMetGlnGlyAspPheTyrThrLysGly 223
DB 49365 GCGTTCGCCAAGTGTGACACGACGATCACTGATGCAAGGCGGATTTCTATACACCGGC 49306
QY 224 LysTyrGlyGlnGlyLeuGlnProPheAspMetGluLysAlaIleArgGluAspAla 243
DB 49305 AAATCCCGGAGAAAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 49246
QY 244 GluTyrValValPheAsnGlySerValGlyAlaLeuThrGlyGluAsnAlaLeuLysAla 263
DB 49245 ACCTACGTCCTCTTCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 49186
QY 264 LysValGlyGluThrValArgLeuPheValGlyAsnGlyGlyProAsnLeuThrSerSer 283
DB 49185 AAGACCGGAGACGCGTGTGATCTTCGCGCAACGCGGCGCGCGCGCGCGCGCGCGCG 49126
QY 284 PheHisValIleGlyGluIlePheAspLysValHisPheGluGlyGlyLysGluAsn 303
DB 49125 TTCACGATGCGCGCGCATCTTCATGACGTGGCGGCGCGCGCGCGCGCGCGCGCGCG 49066
QY 304 HisAsnIleGlnThrThrLeuIleProAlaGlyGlyAlaIleThrGluPheLysVal 323
DB 49065 AAGATGTGCGACACACCTGATCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 49006
QY 324 AspValProGlyAspTyrValLeuValAsnHisAlaIlePheArgAlaPheAsnLysGly 343
DB 49005 AGGTCGCGCGGAGCTAGCTGTGTGACCACTGATTTTCGCGCTTCAACAGGGG 48946
QY 344 AlaLeuGlyIleLeuLysValGluGluGluAsnHisGluIleTyrSerHisLysGln 363
DB 48945 GCGATGCGCATCTCAAAATGACAGCGTCCGGAAGACAAACCGCTGTATTCGCGCAAGG 48886
QY 364 ThrAspAlaValTyrLeuProGluValA---ProGlnAlaIleAspThrGlnGluAla 382
DB 48885 CTGATGCGGTGTACTGTGGCGACCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 48826
QY 383 ProLysThrProAlaProAlaAsnLeu-----GlnGluGlnIleLysAlaLysAla 400
DB 48825 GCGCAGGCGTCCGCTCCGCGACCGCTGTGATGACGATGACGATGACGACGCGCGCGG 48766
QY 401 ThrTyrAspSerAsnCySerAlaAlaCyHisGlnProAspGlyLysGlyValProAsnAla 420
DB 48765 CTGTTGCGGGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 48706
QY 421 PheProProLeuAlaAsnSerAspTyrLeuAsnAlaAspHisAlaArgAlaAlaSerIle 440
DB 48705 TTTCGCGCGCTGGCCAAATCGGACTTCTGGCGGCGGCGCGCGCGCGCGCGCGCGCG 48646
QY 441 ValAlaAsnGlyLeuSerGlyLysIleThrValAsnGlyAsnGlnTyrGluSerValMet 460
DB 48645 GTGTCGACGCGCTGAACGCGCAAGTCAAGGTAATGCGCAGAAATACGATTCGCTGATG 48586

```

Oy		461	Prohalitealae---LenseAsgGlnGlnlellaenavallierTyrThrLeuan	479
Dd		48585	CCCGCGAGACGCAGCTCAACGATGACGAGGTGCCAACACTCCTGACCTTAGTGCTTAAC	485286
Oy		480	SerPhegliaenlygIylGInLeuSerAlaAspApValAlaYsaLatylSyrthr	499
Dd		48525	AGCTGGGACAACC GG GGG CGCGGT CAGCGCGGAAGACGTC---AAGACGATTCTGTC	48463
Oy		500	LysPro 501	
Dd		48468	CAGCCC 48463	
RESULT 6				
AEO02512				
LOCUS				
DEFINITION				
VERSION				
KEYWORDS				
ORGANISM				
REFERENCE				
AUTHORS				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				

gene  
CDS  
AONTLVGAVALVPTPETTTAOTIOAIRREDFIRAKIGMLGPAEIIIECVADKLKCSG  
RRAVDLPWAIAGAPKLQDSVAALTRILLPDTVLPUNLPEAALGVHNRKNDK  
RAKLLIVGVANVIRKGLHSTSGCTWMLTFQNETLEFDSPPRTATHTGCTG  
FSACTAIABLKQSDCEAVQTAKEYITTAISNPLEIGAGHPVNMAYRD"  
complement (1209) .2078)  
/gene="NMB1617"  
/complement (1209) .2078)  
/gene="NMB1617"  
/note="similar to GB:M74072 SP:P25397 PID:149018 GB:U00966  
PID:1742373;percent identity: 68.42; identified by  
sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="cellulurite resistance protein, putative"  
/protein\_id="AAFA1969.1"  
/db\_xref="GI:7226868"  
/translation="WKERIVGQSGELPCFGQMPVAKYENTPEVLLSGYSSEEGEWCLL  
NYLQDSVVRAPADSEAVSWASGDGVAPQGVSPVKPTQDAIRLSLYCAADYPRH  
KYGMSADHVSVAAPARDTVAPAGRALDMQCGGRNALPLGLKGFTVTRVDRNPALAVG  
AELABEGIANVRLTEYDLMNALGSEPDYIATVLLPLMFQRPVDYIADQATTAAG  
GNLVLVSMADPAPFCPPMPFPFKREGELADKYRDMELVEYKEELGSMHAKDENGPI  
RKFPTMLAKRPE"  
2192 .2629  
/gene="NMB1618"  
2192 .2629  
/gene="NMB1618"  
/note="similar to SP:P00647 PID:1208974 PID:147677  
PID:147680 PID:42062 percent identity: 73.94; identified  
by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="ribonuclease H1"  
/protein\_id="AAFA1970.1"  
/db\_xref="GI:7226869"  
/translation="MNQTVLYTDGACKGNPGAGMGVLMRYSHEKELFGSEKQNTN  
NRMETAVITEGLKRCCTVITCTDSQYVNGKEMNIHGMKRGMTAKRQPKYNDN  
LWKELDALVGHQVSWTVKVGKAGHAENERRADLIANGAAGFS"  
3487 .3891  
/gene="NMB1619"  
3487 .3891  
/gene="NMB1619"  
/note="conserved hypothetical protein; identified by  
Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAFA1971.1"  
/db\_xref="GI:7226870"  
/translation="WTAQPCPTCAQMEDVILQTPNLRVAVANDSSPAFCRYTRK  
HIAETPDISAERGELMEWYKVAAMRQVFRPAKINLASLGNVPHLWHIIARFEN  
DASFPAPITWAPVVRKHGWTLPQDWTEQLKLL"  
3931 .5115  
/gene="NMB1620"  
3931 .5115  
/gene="NMB1620"  
/note="conserved hypothetical protein; identified by  
Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAFA1972.1"  
/db\_xref="GI:7226871"  
/translation="MTPEPTSRRRFLTKTCTAAAGAGLLQAGTSATSVPLPSHSYVK  
ARVYLPQRRQSSDGNILRVASGPAEDTNRYNTALTLYNNGFVTVQAGSRRE  
ORPAGTDOPRAQROEVAAGRVATPKVLMGLRGYGAARILPHIDPASIGABMRHEQT  
LPRGSDYCAVOLALLAKGNMNSPAGPMAYSDRKPARGAFPMDAFIKGTQNRUTYD  
VPIYQRADEVTEGILMGJNLSVLASLGTPEMDIDGILPLEDVGSQPIREIMLNT  
LYSGILDKQCAIVGVNFRMRKIRADVDPSEYSAVANNVSRAKIPLVLFQFPGHIA  
DKITFIPGAHRIIRMGNSGYSAFEGYPTLIDASALTLDTLPPDPPIIPESGVADI  
SE"  
complement (5178) .5711)  
/gene="NMB1621"  
complement (5178) .5711)  
CDS



```
/gene="NM1621"
/notes="similar to PID:642472 SP:P52036 percent identity:
100.00; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="glutathione peroxidase"
/protein_id="AAFA1973.1"
/db_xref="GI:7226872"
/translation="MGIYDFQMKDAEGNAVDLSGKGVLLIVNTATRCGLTPQYEA
LQVLAQVTAEGEILDFPCNOREQAPESSEGIAYCQMKFGTKFKIPKIRVNGANT
APLVAIKSKVKPKDGNHFKDFVLKLAALGKRDGDIKMMFTKFLVNRDSEVERF
ASVTPREIEADIRALL"
/complement(5890..8145)
/gene="NM1622"
/complement(5890..8145)
/notes="similar to GP:2584871 percent identity: 78.28;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="nitric oxide reductase"
/protein_id="AAFA1974.1"
/db_xref="GI:7226873"
/translation="MGQYKLMYLPVAVLAVGFTLIGNGSEVYKKAAPPYPROVYSAS
GVVLTAKDILAGQSAWTTGMEVSLIGHAYQAPDWTALMLHRELVAMLDLTAQ
TYGKKFEVSPBEOALNTRLADENYNOGRVSDGSVLSIDRVKAIIESILPYHGVY
SDPKFOETREHAPMKNLTPSOAREKLFNFEPMTWSASTNRPEENTYNNMHE
PLINNVPTENTMMSFTSVLLMGLGILMGLMGLYFTKIEEVEPESDPSIKIOLTPS
OKALGVYFLVALPVQVYLIGLTAHYVGGGPGFIDGAEADPEMSDPALITRP
HIQSAFWIATGFTLNGFLAPYVNGKPKFORAGVNTLYALFTVVGSGTAGNFA
LTHLPBEPFNGHOGYBYLDGRFWQLLVGLMLFMLRCTVSAFKERGVKN
LAIATVAVNGVGFYAPGLFYGEKSPVAVMEYRWVVALVEGFEPVAPAAFAFV
FVNGVFRSTASTASTLAAALFMLGSGVTLHLVYFSGTSASMAIGAFSLEVP
LVLLRSEAYEHSYHSDMKARLWPLMCFVAVAPMNIAGVFGFLINPISLPIYI
OGLNLSAVAHALRGVGFALGFVLVARYLKPQADDDLMTMGFWLLNGLAGM
IAISLIPVGAIQAYASITGTLNVARSEEPLOMEILDITLNMVTRADLVITGCAICV
AIQATVFGDK"
/gene="NM1623"
/516..9688
/8516..9688
/notes="similar to GB:M97926 SP:Q02219 PID:150277 percent
identity: 98.72; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=11
/product="major anaerobically induced outer membrane
protein"
/protein_id="AAFA1975.1"
/db_xref="GI:7226874"
/translation="MKRQALAAIASLFPALACGEPAAQAPETPPAAAASAAQ
TAETPSGELPVIDATTHAPVPALIDDPYAKVRKMETVBEKMTMDGDEYRWT
FGDVPBPMIRBEDGTVEFSNNSSVPNNVDPHATGGCGAAALFTAPGRST
FSFKALQPGLYTHCAVAFVGMHIANGMGLILVEPKGLPYDKEFYIVQDFTTKG
KKGAQGLPDMDKAAEQPEYVFNHGAIGALADALAKAGETVRMVGNGGPLYV
SSEFHIGELFDKVVYEGKLINENQSTIVPAGSAIVFKVDIPESYTLVHSHIFRA
FKKALGOLKVGAEENPEIMTKLSDTAVANGAAPAAAPASAPASASEKSYV"
BASE COUNT 2477 a 2767 c 2385 g 2163 t
ORIGIN
```

Alignment Scores:

Pred. No.:	2,02e-83	Length:	9792
Score:	1394.00	Matches:	265
Percent Similarity:	79.74%	Conservative:	46
Best Local Similarity:	67.95%	Mismatches:	75
Query Match:	53.08%	Indels:	4
DB:	1	Gaps:	2

US-10-088-045-2 (1-502) x AE002512 (1-9792)

QY 1 MetSerLysProThrLeuIleLysThrThrLeuIleCysAlaLeuSerAlaLeuMetLeu 20

DB 8498 TTACAAAGGAAATATTTATGAAACCGCAAGCCTTAGCTGCAATGATGCTTCCTTATTC 8557

```
QY 21 SerGlyCysSerAsnGlnAlaAspLysAlaGlnProLysSerSerThrValAspAla 40
DB 8558 GCATTAGCCGCTCGCGCGCGCGAACCCTGCCGCAAGCCCTGCCGAAACCCCTGCCCT 8617
QY 41 AlaAlaLysThrAla---AsnAlaAspAsnAlaAlaSerGlnGlnLHisGlnLysGlnLeu 59
DB 8618 GCGCGCGAAGCCGCAAGCTCCGCCACAAACCGCGCGGAAACACCGCTCGCGCAACG 8677
QY 60 ProValIleAspAlaIleValThrHisAlaProGluValProProProValAspArgAsp 79
DB 8678 CCCGTTATCGATGCGGTTACCAACCGCTCCGCAAGTCCCTCTCGCATCAACCGCGAC 8737
QY 80 HisProAlaLysValValValLysMetGlnThrValGluLysValMetArgLeuAlaAsp 99
DB 8738 TACCCGCGCAAGTCCGCGTAAATGGAACCGTCGAAAAAACATGACCATGGAAGAC 8797
QY 100 GlyValGluTyrGlnPheThrThrPheGlyGlyGlnValProGlyGlnMetIleArgVal 119
DB 8798 GGTGTGGAATACCGGTACTGACATTTGACGCGGACGTCCTCCGCGGTATGATCCGCGTA 8857
QY 120 ArgGluGlyAspThrIleGluValGlnPheSerAsnHisProAspSerLysMetProHis 139
DB 8858 CCGGAAGCGATACCGGTGAAGTGAATTTCCAAACATCTCTTACCGTCCGAC 8917
QY 140 AsnValAspPheHisAlaAlaThrGlyProGlyGlyGlyAlaGlnLysPheThrAla 159
DB 8918 AACGTGACATTCACGCGGCTACCGGCGGCGGCGGCGGCGGCGCAACCTTTACCGCT 8977
QY 160 ProGlyHisThrSerThrPheSerPheLysAlaLeuGlnProGlyLysThrValTyrHis 179
DB 8978 CCGGCGCGTACTTCACTTCACTTCACTTCAAGCCCTGCAACCGGCTGTACTTCAAC 9037
QY 180 CysAlaValAlaProValGlyMetHisIleAlaAsnGlyMetTyrGlyLeuIleLeuVal 199
DB 9038 TCGCGCGTCCGACCGGTCGATGCAATCGCCACCGTATGATGATGATTTTGGTC 9097
QY 200 GluProLysGluGlyLysProLysValAspLysGluTyrTyrValMetGlnLysAspHe 219
DB 9098 GAGCTTAAAGAGCGCTCGCGAAGTGAATGATTAAGGTTCTATCATGCTCAAGCGCATTC 9157
QY 220 TyrThrLysGlyLysTyrGlyGlnGlnLysLeuGlnProPheAspMetGlyLysAlaIle 239
DB 9158 TACACCAAGGCAAAAGGCGCGCAAGTCTGCAACCGTTCGATATGACCAAGCGCTT 9217
QY 240 ArgGluAspAlaGluTyrValValPheAsnGlySerValGlyAlaLeuThrGlyGluAsn 259
DB 9218 GCGGAACGCTGATATGATGATTCACCGTCAAGTGTGCTATGCGCGCGATAC 9277
QY 260 AlaLeuLysAlaLysValGlyGluThrValArgLeuPheValGlyLysAsnGlyLysProAsn 279
DB 9278 GCGCTGAAAGCAAGCAAGCGCAAGCTGATGATGATGATGATGATGATGATGATGATGAT 9337
QY 280 LeuThrSerSerPheHisValIleGlyGluLysPheAspLysValHisPheGluGlyGly 299
DB 9338 TTGGATATCTTCTCTTCAAGTCTTCAAGTCTTCAAGTCTTCAAGTCTTCAAGTCTTCA 9397
QY 300 LysGlyGluAsnHisAsnIleGlnThrThrLeuIleProAlaGlyAlaAlaIleThr 319
DB 9398 AACTGATTTAAGAAAGATGATCAAGCAACATGCTTCTCGCGCGGCTCTCCATCGTC 9457
QY 320 GluPheLysValAspValProGlyAspTyrValLysValAspHisAlaIlePheArgAla 339
DB 9458 GAATTCAAAGTGCATCCCGGCGAGTCACTTGTGTTGACCATCTATCTTCCGCGCA 9517
QY 340 PheAsnLysGlyAlaLeuGlyLysLeuLysValGluGlyGluLysAsnHisGluLysTyr 359
DB 9518 TTCAACAAAGGCGCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 9577
QY 360 SerHisLysGlnThrAspAlaValTyrLeuProGluGlyAlaProGlnAlaIleAspThr 379
DB 9578 ACTCAAAATGATGATACCGCTTACCGCGGTAACGGTGCAGCTCTGCT----- 9628
```

OY 380 GlnGluAlaProlyThrProAlaProAla 389  
 Db 9629 GCTTCCGCTCCCGACGCTTCTGCCCGCA 9658  
 RESULT 7  
 NMA622491 326301 bp DNA linear BCT 02-SEP-2002  
 LOCUS Neisseria meningitidis serogroup A strain Z2491 complete genome;  
 DEFINITION segment 6/7.  
 ACCESSION AL162757 AL157959  
 VERSION AL162757.2 GI:7380371  
 KEYWORDS  
 SOURCE Neisseria meningitidis Z2491  
 ORGANISM Neisseria meningitidis Z2491  
 Neisseria: Proteobacteria; Betaproteobacteria; Neisseriales;  
 Neisseriaceae; Neisseria.  
 REFERENCE 1 (bases 1 to 326301)  
 AUTHORS Parkhill,J., Achtman,M., James,K.D., Bentley,S.D., Churcher,C.,  
 Klee,S.R., Morelli,G., Basham,D., Brown,D., Chillingworth,T.,  
 Davies,R.M., Davis,P., Devlin,K., Feltwell,T., Hamlin,N.,  
 Holroyd,S., Jagels,K., Leather,S., Moule,S., Mungall,K.,  
 Quail,M.A., Rajandream,M.A., Rutherford,K.M., Simmonds,M.,  
 Skelton,J., Whitehead,S., Spratt,B.G. and Barrall,B.G.  
 TITLE Complete DNA sequence of a serogroup A strain of Neisseria  
 meningitidis Z2491  
 JOURNAL Nature 404 (6777), 502-506 (2000)  
 MEDLINE 20222556  
 PUBMED 10761919  
 REFERENCE 2 (bases 1 to 326301)  
 AUTHORS Parkhill,J.  
 TITLE Direct Submission  
 JOURNAL Submitted (30-MAR-2000) Submitted on behalf of the Neisseria  
 sequencing team, Sanger Centre, Wellcome Trust Genome Campus,  
 Hinxton, Cambridgeshire CB10 1SA E-mail: parkhill@sanger.ac.uk  
 COMMENT Details of N. meningitidis sequencing at the Sanger Centre are  
 available on the World Wide Web.  
 (URL: [http://www.sanger.ac.uk/Projects/N\\_meningitidis/](http://www.sanger.ac.uk/Projects/N_meningitidis/)).  
 FEATURES  
 source 1. 326301  
 /organism="Neisseria meningitidis Z2491"  
 /mol\_type="genomic DNA"  
 /strain="Z2491"  
 /db\_xref="taxon:122587"  
 /note="serogroup: A"  
 196. .311  
 /note="REP 2, hmms hit to HMM REP 2 (1 - 109), score:  
 87.43"  
 434. .453  
 /note=">= 90% match to ATCCGNNNNNNNGGAAT"  
 /label="dRS3"  
 498. .560  
 /note="Correia element; hmms hit to HMM Correia (1 - 62),  
 score: 78.71"  
 /label="Correia"  
 562. .604  
 /note="Correia element; hmms hit to HMM Correia (114 -  
 156), score: 51.92"  
 /label="Correia"  
 complement (667. .676)  
 /note="Core DNA uptake sequence: gccgcctgaa"  
 /label=DUS  
 complement (668. .1813)  
 /gene="dape"  
 complement (668. .1813)  
 /gene="dape"  
 complement (668. .1813)  
 /gene="dape"  
 /EC\_number="3.5.1.18"  
 /note="NMA1730, dape, succinyl-diaminopimelate  
 desuccinylase, len: 381aa; similar to many eg. SW:P24176  
 (DAPE\_ECOLI) dape, succinyl-diaminopimelate desuccinylase  
 from Escherichia coli (375 aa) fasta scores; E(): 0, 56.9%  
 identity in 371 aa overlap. Contains Pfam match to entry  
 PF01546 Peptidase\_M20."

misc\_feature  
 /codon\_start=1  
 /transl\_table=11  
 /product="succinyl-diaminopimelate desuccinylase"  
 /protein\_id="CAB84958.1"  
 /db\_xref="GI:7380372"  
 /db\_xref="SPTREMBL:Q9JTL0"  
 /translation="MTETQSLLEARALISRPSTVTPDDRCCQKLVERLYKIFGADEL  
 HFGDTRKNIWLRGTGVVPCFAGHDVPTGVKWDSPFETRDGLYRGADAM  
 KTIACFPVACERFVAEHPDHQGSIALITSPEDADLDTTKVVDYKARBELDYC  
 IYGEPRADKGDGMKNGRGSLSNLTAVKQGHIAVHLAINVHTAPALLETQ  
 EYWDKNTKFPPTSTFOISVINGTGATVNIPEELVKNENRFTSTETAGLQKHAI  
 LDKHGVQYDLQWSCSGQPLTQAGKLTVAARAALAEIGIEALISTGTSGDRPTKA  
 IAKELIELGPSNATTHQINENVRDLDPRLSAVYEGIIARLIGNAV"  
 complement (836. .1786)  
 /gene="dape"  
 /note="Pfam match to entry PF01546 Peptidase\_M20, score  
 252.50, E-value 5.7e-72"  
 complement (1863. .1872)  
 /note="Core DNA uptake sequence: gccgcctgaa"  
 /label=DUS  
 complement (1927. .2547)  
 /gene="NMA1731"  
 complement (1927. .2547)  
 /gene="NMA1731"  
 /note="NMA1731, conserved hypothetical protein, len:  
 206aa; similar to SW:P44013 (Y552 HAEIN) hypothetical  
 protein from Haemophilus influenzae (207 aa) fasta scores;  
 E(): 0, 59.6% identity in 203 aa overlap."  
 /codon\_start=1  
 /transl\_table=11  
 /product="conserved hypothetical protein"  
 /protein\_id="CAB84959.1"  
 /db\_xref="GI:7380373"  
 /db\_xref="SPTREMBL:Q9JQZ3"  
 /translation="WLTPKSCDLENIIPFPQSOLKKYPPESIPQIKADYKNWQIQWQ  
 LIQVAAELGAFAPAPHIERMCGVAVHFAFYKQYKNSAAILSLILNRRLSV  
 LIDMHCYKADVSPIALPDYRNLDNFDYKASFPMMGAESEYDYDTVAQONESDR  
 KRONEDPFCIGKIERDLDGRODVAKVIAETVEDLLPLYERCHGK"  
 complement (2551. .2558)  
 complement (2570. .2579)  
 /note="Core DNA uptake sequence: gccgcctgaa"  
 /label=DUS  
 complement (2581. .3090)  
 /gene="NMA1732"  
 complement (2581. .3090)  
 /gene="NMA1732"  
 /note="NMA1732, conserved hypothetical protein, len:  
 169aa; similar to SW:P43971 (Y243 HAEIN) hypothetical  
 protein from Haemophilus influenzae (172 aa) fasta scores;  
 E(): 0, 49.7% identity in 169 aa overlap."  
 /codon\_start=1  
 /transl\_table=11  
 /product="conserved hypothetical protein"  
 /protein\_id="CAB84960.1"  
 /db\_xref="GI:7380374"  
 /db\_xref="SPTREMBL:Q9JTK9"  
 /translation="MNPETOSVTPAEIEMLVACHGKVRPPCGQAWLSQYIAENG  
 NQVIVQTRQITRYENVAPLPHHEBENFPPLIQVAPQAGSDDELRLQVSLHNG  
 WDAVAERPAKLQADNAVTPDAEAFRRFVAGVHVALEPPLDWMKMTPEKLTETG  
 EIMARRRRR"  
 complement (3185. .3736)  
 /gene="NMA1733"  
 complement (3185. .3736)  
 /gene="NMA1733"  
 /note="NMA1733, H.8 outer membrane protein, len: 183aa;  
 strongly similar to SW:P07212 (H8 NEIME) H.8 outer  
 membrane protein from Neisseria meningitidis (183 aa)  
 fasta scores; E(): 0, 96.7% identity in 183 aa overlap.  
 Contains Pfam match to entry PF00127 copper-bind. Copper  
 binding proteins, plastocyanin/azurin family and. Prosite  
 match to PS00196 Type-I copper (blue) proteins signature."  
 /codon\_start=1  
 /transl\_table=11



QY 381 GluAlaProLysThrProAlaProAla 389  
 Db 149109 TCCGCTCCCGACCTCTGCTCTGCA 149135

RESULT 8  
 LOCUS AX044033 349980 bp DNA linear PAT 24-NOV-2000  
 DEFINITION Sequence 112 from Patent WO0066791.  
 ACCESSION AX044033  
 VERSION AX044033.1 GI:11342917  
 KEYWORDS  
 SOURCE Neisseria meningitidis  
 ORGANISM Neisseria meningitidis  
 Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales; Neisseriaceae; Neisseria.

REFERENCE  
 AUTHORS Pizzo, M., Hickey, E., Peterson, J., Tetteijn, H., Venter, J. C., Masignani, V., Galeotti, C., Mora, M., Ratti, G., Scarcell, M., Scariato, V., Rappunli, R., Frazer, C. M. and Grandi, G.  
 TITLE Neisseria genomic sequences and methods of their use  
 JOURNAL Patent: WO 0066791-A 112 09-NOV-2000;  
 CHIRON CORPORATION (US); THE INSTITUTE FOR GENOMIC RESEARCH (US)

FEATURES  
 source  
 1. 349980  
 Location/Qualifiers  
 /organism="Neisseria meningitidis"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:487"  
 /note="sequence too long, cut in 8 pieces.-seq 1: 1 to 349980 349980 bases-seq 108: 300001 to 649980 349980 bases-seq 109: 600001 to 949980 349980 bases-seq 110: 900001 to 1249980 349980 bases-seq 111: 1200001 to 1549980 349980 bases-seq 112: 1500001 to 1849980 349980 bases-seq 113: 1800001 to 2149980 349980 bases-seq 114: 2100001 to 222325 172325 bases"

BASE COUNT 87189 a 93501 c 84627 g 84663 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 1,19e-81 Length: 349980  
 Score: 1394.00 Matches: 265  
 Percent Similarity: 79.74% Conservative: 46  
 Best Local Similarity: 67.95% Mismatches: 75  
 Query Match: 53.08% Indels: 4  
 Gaps: 2

US-10-088-045-2 (1-502) x AX044033 (1-349980)

QY 1 MetSerLysEProThrLeuLysThrThrLeuLysCysAlaLeuSerAlaLeuMetLeu 20  
 Db 187900 TTCACAAAGGAAATATTATGAAAGCGCAAGCCTTAGCTGCAATGATGCTTCTTATTC 187959

QY 21 SerGlyCysSerSerGlnAlaLeuLysAlaGlnProLysSerThrValaAla 40  
 Db 187960 GCATTAGCCGCGCGCGCGCAACTGCGCGCAAGCCCTCCCAACCCCTGCGCGCT 188019

QY 41 AlaAlaLysThrAla--AenAlaAspAlaAlaLysSerGlnLysGlnLysGlnLysGlnLys 59  
 Db 188020 GCCCGCAAGCCGCAAGCTCCGCGCAACACCGCGCAACACCGCGCGCAACCTG 188079

QY 60 ProValIleAspAlaIleValIleThrAlaProGluValProProProValaAsp 79  
 Db 188080 CCGGTTATCGATCGGTTACCAACGCTCCGGAAGTCTCTGCAATGACCGCGAC 188139

QY 80 HisProAlaLysValaValaValaLysMetGluThrValGluLysValaMetArgLeuAlaAsp 99  
 Db 188140 TACCCCGCAAAAGTCCGCGCAAAAGCAAAACCGTGAAGAAACCAATGACCAAGAAAGAC 188199

QY 100 GlyValaGluTyrGlnPheThrPheGlyGlyGluValProGlyGlnMetIleArgVal 119  
 Db 188200 GGTGTGAATACCGCTACTGACATTTGACGGGACGTTCCGGCGGATGATCCCGCTA 188259

QY 120 ArgGluGlyAspThrIleGluValGlnPheSerAsnHisProAspSerLysMetProHis 139

Db 188260 CCGCAAGCGATACGGTTGAAGTGAATTTCCAAACATCTTCTTCAACCGTCCGAC 188319

QY 140 AsnValaAspPheHisAlaAlaIleThrGlyProGlyGlyValaGluAlaSerPheThrAla 159  
 Db 188320 AACGTGATCTCCACGCGCTCCGCGCAAGCGCGCGCGCGCGCAACCTTACCGCT 188379

QY 160 ProGlyHisEThrSerThrPheSerPheLysAlaLeuGlnProGlyLeuThrValTyrHis 179  
 Db 188380 CCGGCGCGTACTTCCATCTTCACTTCCAAAGCCCTCCCAACCGGCTGTGACATCTACAC 188439

QY 180 CysAlaValaIleProValaGlyMetHisIleAlaAsnGlyMetTyrGlyLeuVal 199  
 Db 188440 TCGCGCGTGCACCGGTGCGATGACATCGCAACGGATATGACGTTGATTTGGTC 188499

QY 200 GluProLysGluLysLeuProLysValaLysGlyTyrTyrValaMetGlnLysAspPhe 219  
 Db 188500 GAGCCTAAAGAAAGGCTGCGCAAGGATGAAGATTTACATGCTCCAAAGCACTTC 188559

QY 220 TyrThrLysGlyLysTyrGlyGluGlnLysGlnProPheAspMetGluLysAlaIle 239  
 Db 188560 TACACCAAGGCAAAAGGCGCGCAAGGCTCTGCAACGTTGATATGACAAAGCCGT 188619

QY 240 ArgGluAspAlaGluTyrValaValaPheAsnGlySerValaGlyAlaLeuThrGlyLys 259  
 Db 188620 GCCGAACGCTGAATAGCTGATTAACAGTCAAGTCAAGTGTGATATGCGCGCGATAC 188679

QY 260 AlaLeuLysAlaLysValaGlyLysThrValaValaGluPheValaGlyAlaAsnGlyProAsn 279  
 Db 188680 GCGCGAAAGCCCAAGCAAGCGCAAACTGATGATGCTGTTGATGACGCGCTCGAAC 188739

QY 280 LeuThrSerSerPheHisEValIleGlyGluIlePheAspLysValaHisPheGluGlyGly 299  
 Db 188740 TTGATATCTTCTTCCACGCTCATCGGCAATCTTCCGCAAGTTATGTTAAAGCGGC 188799

QY 300 LysGlyLysAsnHisAsnIleGlnThrThrLeuLysProAlaGlyValaAlaIleThr 319  
 Db 188800 AAAGTATTAACGAAGACGTAACAGCAACATGCTTCCGCGCGGCTCCATCGTC 188859

QY 320 GluPheLysValaAspValProGlyAspLysValaLeuValaAspHisAlaIlePheThrAla 339  
 Db 188860 GAATTCAAAGTGCATCCCGGCGAGTACACTTGTGTGACCACTATCTTCCGCGCA 188919

QY 340 PheAsnLysGlyAlaLeuGlyLysLeuLysValaGluGlyGluAlaAsnHisGluIleTyr 359  
 Db 188920 TTCACAAAGGCGCACTGGGTGAATGAAAGTGAAGTGCAGAAACCTGAATCATG 188979

QY 360 SerHisLysGlnThrAspAlaValTyrLeuProGluGlyAlaProGlnAlaIleAspThr 379  
 Db 188980 ACTCAAAATATGATGATACCGCTTACCGCGGTACGGTGCAGTCTGCT----- 189030

QY 380 GlnGluAlaProLysThrProAlaProAla 389  
 Db 189031 GCTTCCGCTCCGCAAGCTTTCGCGCGCA 189060

RESULT 9  
 LOCUS NGOPANI 2069 bp DNA linear BCT 26-APR-1993  
 DEFINITION Neisseria gonorrhoeae major anaerobically induced outer membrane protein (PanI) gene, complete cds.  
 ACCESSION M97926  
 VERSION M97926.1 GI:150276  
 KEYWORDS anaerobically induced protein; major outer membrane protein.  
 SOURCE Neisseria gonorrhoeae  
 ORGANISM Neisseria gonorrhoeae  
 Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales; Neisseriaceae; Neisseria.

REFERENCE  
 AUTHORS Hoehn, G. T. and Clark, V. L.  
 TITLE Isolation and nucleotide sequence of the gene (anI) encoding the major anaerobically induced outer membrane protein of Neisseria gonorrhoeae  
 JOURNAL Infect. Immun. 60 (11), 4695-4703 (1992)













[illegible]

```

RESULT 14
AB013078
LOCUS       AB013078
DEFINITION  AB013078 2375 bp DNA linear ECT 01-FEB-2000
             Alcaligenes xylosoxidans nir gene for Nitrite Reductase (Nir) ,
             complete cds.
ACCESSION   AB013078
VERSION     AB013078.1 GI:3721763
KEYWORDS    nir; Nitrite Reductase.
SOURCE      Achromobacter xylosoxidans
            Achromobacter xylosoxidans
            Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
            Alcaligenaceae; Achromobacter.
ORGANISM    1 (sites)
            Karsaka,K., Furusawa,H., Yamaguchi,K. and Suzuki,S.
REFERENCE   Cloning and Expression of Copper Nitrite Reductase Gene from
            Alcaligenes xylosoxidans G1FU1051
            Unpublished
            2 (bases 1 to 2375)
            Suzuki,S. and Karsaka,K.
AUTHORS     Direct Submission
            Submitted (15-APR-1998) Shimichiyo Suzuki, Osaka University,
            Department of Chemistry, Graduate School of Science, Machikaneyama
            1-16, Toyonaka, Osaka 560-0043, Japan
            (E-mail:bioc@ch.wani.osaka-u.ac.jp, Tel:+81-6-850-5768,
            Fax:+81-6-850-5785)
FEATURES
     source
         1..2375
             Location/Qualifiers
                 1..2375
                     /organism="Achromobacter xylosoxidans"
                     /mol_type="genomic DNA"
                     /strain="G1FU1051"
                     /db_xref="taxon:85698"
                     590..602
                         /note="fmr-box"
                     822..1904
                         /gene="nir"
                     822..1904
                         /gene="nir"
                     /codon_start=1
                     /transl_table=1
                     /product="Nitrite Reductase (Nir)"
                     /protein_id="BAJ3678.1"
     cds
         822..1904
             /gene="nir"
             /codon_start=1
             /transl_table=1
             /product="Nitrite Reductase (Nir)"
             /protein_id="BAJ3678.1"

```

[illegible]

Qy	260	AlaLeuValAlaValAlaGlyValThrValArgLeuPheValGlyAsnGlyValPro	279
Db	1569	GGCTCACCCGCCAGAGTGGCGAGACCGTCTGCTGATC-----CATCCGAGCCAA	1622
Qy	280	LeuThrSerSerPheHisValIleGlyValIlePheAspLysValHisPheGluGly	299
Db	1623	CGCGACACCCCGCACCTGATCGGGCGGCATGGCCATGGAGTGGGTC---TGGAGACCGGC	1679
Qy	300	LysGlyGluAsn-----HisAsnIleGlnThrThrLeuIleProAlaGly-----	314
Db	1680	AAGTTGCGCAACCCGCCAGCGCCAGCCATCGAAMACCTGTTATCCCGCGGGGTGGCC	1739
Qy	315	GlyAlaAlaIleThrGluPheLysValAspValProGlyAspTyrValLeuValAspHis	334
Db	1740	GGCGCGGGCTCTACACCTTCAAG-----CAGCTTGGCGTATGCTTACTTACCTGAACCA	1793
Qy	335	AlaIlePheArgAlaPheAsnLysGlyAlaLeuGlyIleLeuLysValGluGlyGlu	354
Db	1794	AACCTGATCGAGCGCTTGCAGACTGGCGCGCGGCCCAATCAAGAGTGAAGGGCAAGTG	1853
Qy	355	AsnHisGluIleTyr-SerHisLysGlnThrAsp-----	365
Db	1854	AACGACGACCTGATGAACAGATCAAGCGCGCCCGCATTCGCGCTGACAGAGCC	1913
Qy	366	-----AlaValTyrLeuProGluGlyValAlaProGlnAlaIleAspThrGlnGly	381
Db	1914	CGCCATGCTCCCTCGCATCAATGCATTTGGCACTGGCGCGCGCTGCGCGCGGGCTGC	1973
Qy	381	ValAlaProLysThrProAlaProAlaAsnLeuGlnGluGlnIleLysAlaGlyLysAlaTh	401
Db	1974	CCATGCGCGCGCGCCGACGCGCGCGCGCATCTGCAC-----CCGCGCGGTGAAMAGCT	2024
Qy	401	RTYrAspSerAsnGlyAlaAlaGlyHisGlnProAspGlyLysGlyValAlaProAlaAlaPh	421
Db	2025	GTACCGCTCGGCTGTGTGTGATGCGAC-----GCCAGCGCGTGGGCCAACGCG--	2073
Qy	421	eProProLeuAlaAsnSerAspTyrLeuAsnAlaAspHisAlaArgAlaAla---SerI	440
Db	2074	-CCCAAGCTGGGCCGACAGACAGCGCTGGGCGCGCTTCTGTGGCGCAGGCGCCGACGCGCT	2132
Qy	440	ValAlaAsnGlyLeuSerGlyLys 448	
Db	2133	GCTGGCCACGCTGCTCAAGGGCAAG 2157	
RESULT 15			
RSU62291			
LOCUS	RSU62291	1307 bp	DNA linear BCT 08-AUG-1996
DEFINITION	Rhodobacter sphaeroides copper containing nitrite reductase (nirK)		
ACCESSION	U62291		
VERSION	U62291.1		
KEYWORDS	GI:1480719		
SOURCE	Rhodobacter sphaeroides		
ORGANISM	Rhodobacter sphaeroides		
	Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales;		
	Rhodobacteraceae; Rhodobacter.		
REFERENCE	1 (bases 1 to 1307)		
AUTHORS	Toques, I.E., Kwiatkowski, A.V., Shi, J. and Shapleigh, J.P.		
TITLE	Characterization and regulation of the gene encoding nitrite		
	reductase in Rhodobacter sphaeroides		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 1307)		
AUTHORS	Toques, I.E., Shi, J. and Shapleigh, J.P.		
TITLE	Direct Submission		
JOURNAL	Submitted (26-JUN-1996) Microbiology, Cornell University, wing		
	Hall, Ithaca, NY 14853-8101, USA		
FEATURES	Location/Qualifiers		
source	1..1307		
	/organism="Rhodobacter sphaeroides"		
	/mol_type="genomic DNA"		
	/strain="2.4.3"		
	/db_xref="taxon:1063"		

protein_bind	56..69
gene	/note="similar to previous fnr binding sequence"
CDS	/bound_moiety="Fnr" /evidence=not_experimental 136..1260 /gene="nirK" 136..1260 /gene="nirK" /codon_start=1 /transl_table=1 /product="copper containing nitrite reductase" /protein_id="AAB05767.1" /db_xref="GI:1480720" /translation="MFMRAALVGAALASAPIVTPAGAOEAPAOLASAAPDLISL PVKHTLVPPRFAHAEQVAASCPVINEFMRIIEKVEQLDEDAVLDAMTFDGSIRP LMTHEEDVYELTLINPENTMPNIDFHAATGALGGGLTLINPGEKVYLRPATPA GAFYHCAFGPMIPMHVVGMAICWILPRDGLKHGKRPVRYDTYVIGESDHYIP KDEBGTYMRSTSEGEYEDMAVMDTLIPSHIVENGAVGALTGGAALKAVGVNIVFV HSQKRDSPHLIGGHGLDVMERGKFPNAPERDLETWFIGGVTAGALYKFLQGVYA YVNHLILEAHVKATASVLVGEEMNDLMDQVMPVGLTG"
BASE COUNT	237 a 472 c 405 g 193 t
ORIGIN	

Alignment Scores:		
Pred. No.:	5,93e-23	Length: 1307
Score:	474.00	Matches: 138
Percent Similarity:	46.83%	Conservative: 54
Best Local Similarity:	33.66%	Mismatches: 135
Query Match:	18.05%	Indels: 83
DB:	1	Gaps: 16

QY 34 .....LysThrAla 45  
 Db 148 CGCGCGCGCTCGGGAGCGCGCGCGCTGCGTCAAGCGCGCTGCATCGGAGCGCA 207  
 QY 46 AsnAlaSerAsnAlaIleSerGlnGlnHISglnGlyGluLeuProValIleAspAlaIle 65  
 Db 208 GCGCGCGCAGAGAGCTCGCGCCGCG-----CTT 234  
 QY 66 ValThrHisAlaProGlu-----ValPro 73  
 Db 225 GCGAGTGAAGCGCGCGCTGCATCTCAGCAACCTGCGCGCGTGAAGACACAGCTGCTCT 294  
 QY 74 ProProValAspArgAspHis-----ProAlaIysValVal 86  
 Db 295 CGCGCGCTTCGCGCATGCCCAAGCAGAGTGGCGCGCGCGCGCGCTGCATCAAGCAATTC 354  
 QY 87 LysMetGluThrValGlnIysValMetArgLeuAlaAspGlyValGlnIyrGlnPheTyr 106  
 Db 355 GAGATCGCGATCATCGAGAGAGAAAGTGCAGCTGACAGCAAGATGCTTACTTCGACGGCATG 414  
 QY 107 ThrPheGlnGlyGlnValProGlyGlnMetIleArgValIArgGlnGlyAspThrIleGlu 126  
 Db 415 ACCTTCGACGCGCTGCATCCCGCGCGCGCTGATGATGTCGATGAAGGCGCATGATGCGAA 474  
 QY 127 ValGlnPheSerAsnHisIProAspSerIysMetProHISAsnValAspPheHisAlaAla 146  
 Db 475 CTCACCTGCATCAACCGCGCGCAGAAACACATGCCCCACAAACGACTTCCAGCGCGCC 534  
 QY 147 ThrGlyProGlnGlyGlnValGlnAlaIleSerPheThrAlaProGlnHisIsthrSerThrPhe 166  
 Db 535 ACCGCGCGCGCTGGAGGCGCGGCGCTACGCTCATCAATCCGGCGGAAAGTGTCTCTG 594  
 QY 167 SerPheIysAlaLeuGlnProGlyLeuTyrValIyrHisCysAlaValAlaIleProValGly 186  
 Db 595 CGGTTCAAGGCGCAGCGCGCGCGCGCTGTGTCATCATGCG-----GCCCGCGCGCG 648  
 QY 187 -----MetHisIleAlaAsnGlyMetIyrGlyLeuIleLeuValGlnProIys 202  
 Db 649 CCGATGATCCCTTGACAGTCGTCCTCGGAGATGGCGGCGCTGCATCAAGTGTGTCGCGCG 708

QY	LOCUS	ACNTTRD	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	MEDLINE	PUBMED	REFERENCE	TITLE	FEATURES	FEATURES
QY	203	glnGlyLeuPro-----	lysValAspLysIleuTyrValMet	215	.....													
Db	709	GACGGCTTAAGAACCAACAGAGGCAAGCCGGTCGATTATGACACCGCTATTATACATCGGC	768	.....														
QY	216	glnGlyAspPheTyr-----	ThrTsgLysTyrGlyGluGlnIleuGln	231	.....													
Db	769	GAGAGGACCACTATCATCCCGAAGGACGAGAACCGGACCTAC-----	ATG	813	.....													
QY	232	ProPheAspMetCyluLysAlaIleArgGluAsp-----	-----	242	.....													
Db	814	CGCTTCTCGACCCCGCTCGAGGGCTACGAGGACATGTCGCGGTGTGACACGCTGATC	873	.....														
QY	243	AlaGluTyrValValPheAsnGlySerValGlyAlaIleuThrGlyGluAsnAlaIleuLys	262	.....														
Db	874	CCCAGCCAAATGCTGCTTCAACGGCGGGGGGGCCCTTAACGGCGAGGCCCGGCTGAAG	933	.....														
QY	263	AlaLysValGlyGluThrValArgLeuPheValGlyAsnGlyGlyProAsnLeuThrSer	282	.....														
Db	934	GCCAAAGTGGGGGACAAAGTC---CTCTTCGTC---CAAGCCAGCCCAAGGCCGACTCG	987	.....														
QY	283	SerPheHisValIleGlyGluIlePheAspLysValHisPheGluGlyGlyLysGlyGlu	302	.....														
Db	988	CGCCCGCACCTCATCGGCGGACACGGCGATCTCGTC---TGGGAAACCGGCAAGTTCCAC	1044	.....														
QY	303	Asn-----HisAsnIleGlnThrThrLeuIleProAlaGlyGlyAlaIleThr	319	.....														
Db	1045	AACGGCGCCGACCGTACCTCGACAGCTCGTGTTCACTCCGGCGGACCGGCCGGGCGCG	1100	.....														
QY	320	GluPheLysValAspValProGlyAspTyrValIleValAspHisAlaIlePheArgAla	339	.....														
Db	1105	CTCTACAAAGTTCCTCCAGCCGGCGCTCAAGCGCTACGCTAACCACTATGATCGAGCG	1166	.....														
QY	340	PheAsnLysGlyAlaIleuGlyTyrLeuLysValGlyGlyGluAsnHisGluIleTyr	359	.....														
Db	1165	GTCCACAAAGGGCGCCACGCGCACGCGCTCTCGAGGGCGAATGGGACAAACGACTGG---	1222	.....														
QY	360	SerHisLysGlnThrAspAlaValTyrLeuProGlyGlyAlaProGlnAlaIleAspThr	379	.....														
Db	1222	-----ATGGAAACAGGTGTGGCCGCGGTGCGCTCACCG-----GCTGACCC	1263	.....														
QY	379	glnGluAlaProLysThrProAlaPro	388	.....														
Db	1264	GCCCCGATGCCGCCACACCGACGACGACC	1291	.....														
RESULT 16																		
ACNTTRD																		
LOCUS																		
DEFINITION																		

```

misc_binding      misc_binding      292.307
/bound_moiety="FNR protein"
475.1611
CDS
/codon_start=1
/transl_table=1
/product="precursor of nitrite reductase"
/protein_id="CAA8564.1"
/db_xref="GI:1125639"
/db_xref="SWISS-PROT:P25006"
/translation="MTBQLOMTRRMJAGALAGAVAPLILHTAQAHAGAAAAAGAP
VDISTLPRVAVDVLPFPFAHQOVATQGRVVEFTWTIEKGLVIDRESTEIHMTF
NKSAPGLMIVHENDUYELRLINDPTNLTLHNDPFAATGALGGALTVNPEEETL
RFRTRKGVFVYVHCAPKSGVPMHTVSGMNAIWLIPDGLKDEKGOITFDKIYYGGE
ODPVPVDEAGNKKYETPEGATVEDAYKAMRITLTPHIYVNGAVGALTGHATTAAG
BRVLYVSNQNRDTRPHLIGGHGVYWATCKFNPPLDDETWLIPGTAAGAFYTR
QPGVYAVVNNHLIEAFELGAAGHFXYTGEWDDLMTSVYKPSAM"
sig_peptide
mat_peptide
475.1588
589.1608
/product="nitrite reductase"
BASE COUNT      369 a      601 c      604 g      329 t
ORIGIN

```

Alignment Scores:	
Pred. No.:	1.82e-22
Score:	469.50
Percent Similarity:	44.11%
Best Local Similarity:	32.93%
Query Match:	17.88%
DB:	1
US-10-088-045-2 (1-502) x ACNITRD (1-1903)	
	Length: 1903
	Matches: 162
	Conservative: 55
	Mismatches: 204
	Indels: 72
	Gaps: 17

```

Oy MetLeuSerGlycysSerAenGlnAlaAspGlyAlaGlnProLysSerSerThrVal 38
Db ATGCTTTCGGCGCGCTGCCCTGGCGGC-----GCGGTTCGCGCGCTTCCACACCGCT 561
Oy AspAlaAlaAlaIuSThAlaAsnAlaAspAsnAlaIaSerGlnGlnIuSgInGlyIu 58
Db CAGCGCGATGCGCGTGGCGCTGGCGCGCGGCGAGCGGATGCTCCCGCGATATCTGACC 621
Oy LeuProValAlaIaAspAlaIaIeValThrHisAlaProGluValProProValAspArg 78
Db CTGCGTCCGCTGAAGGTGACATC-----GTCAAGCGCGCTTTCGTGAC 666
Oy AspHis-----ProAlaIuSValValIuSmetGluThrVal 91
Db GCCCATATACAGGTCCGCAAGCCGACCGCGGCTCTCAGATTCAACAAGACATTCGAG 726
Oy GluIuSValIeMetArgLeu--AlaAspGlyValGluIuSglnPheTrpThrPheGlyIu 110
Db GAAAGAAAGCTGGTGATCGACCGCGAGGAGCACTGAGATCCACCGCATGACTTTCACGCT 786
Oy GlnValProGlyGlnMetIeArgValArgGluGlyAspThrIleGluValGlnPheSer 130
Db TCGGTGCTGCTGCTCCGCTGATGGTGCTGATGAGAAGACATATGTTGAACCTGGCGTGATC 846
Oy AsnHisAspAspSerLysMetProHisAsnValAspPheHisAlaIaThrGlyProGly 150
Db AACCTGACACCAACAGCTGTGTCGACCAACATGACTTCCACCGCGAACCGGTGGCGCTC 906
Oy GlyGlyAlaGluAlaSerPheThrAlaProGlyHisIleThrSerThrPheSerPheLysAla 170
Db GCGCGCGGCGCGCTGACCGAGGTGAACCCCGGCGAGAGAAACACAGCTTCTGTTCAAGCGC 966
Oy LeuGlnProGlyLeuIuSValThrHisGlyAlaValAlaIaAspGlyMet-----187
Db ACCAAGCCGCGCGCTTCGTCAATTC-----GCGCGGAGGACAGCGTGCCTGG 1020
Oy HisIleAlaAsnGlyMetTrgGlyLeuIeLeuValGluProGlyGluGlyLeuPro--206
Db CAGCTCACTTCGGCGATGAACGGCGCATATGCTGCTGCGCGGACGAGCATGAAAGAC 1080

```

```

Qy 207 -----LysValAspLysGluTyrTyrValMetGlnGlyAspPheTyr 220
Db 1081 GAGAGGGCCAGCCGCTGACCTGACCAAGATCTACTATGTCGGGACAGACTTCTAC 1140
Qy 221 ThrLys-----GlyTyrTyrGlyGlnGlyLeuGlnProPhe 233
Db 1141 GTCCGGAAGACGAGCCGGGAACTACCAAGAGTACGAAACCCCGGGGAAACCTATGAA 1200
Qy 234 AspMetGluTyrValIleArgGlu---AspAlaGluTyrValIlePheAsnGlySerVal 252
Db 1201 GATCTCTGCAAGGAGATGCGACGCTGACCCCGAACCCATCTCTTCAACGGTGGCTC 1260
Qy 253 G1ValAlaLeuThrGlyGluAsnAlaLeuLysValGlyGluThrValArgLeuPhe 272
Db 1261 GGGCCGGTGAACCGCGGACCATGCTTGTGACTCGCGCGTGGCGAG-----CGTGTCTC 1314
Qy 273 ValGlyAsnGlyGlyProAsnLeuThrSerSerPheHisValIleGlyGluIlePheAsp 292
Db 1315 GTCTCTCATTCGACGAGCCGACCGCGATACGCGCGGACCTGATCGCGGCGATGTCAC 1374
Qy 293 LysValHisPheGluGlyGlyLysGlyGluAsnHis-----AsnIleGlnThrThr 309
Db 1375 TATCTC---TGGGAGACCGCGCAAGTCCGCAACCCCGCATCTGACCAAGAACTCG 1431
Qy 310 LeuIleProAlaGlyGlyValAlaIleThrGluPheLysValAspValProGlyAspTyr 329
Db 1432 CTCACTCCGGCGGACCGCGCGCTGCTTCTTCACTCCGCGACCGCGGTGTGTAC 1491
Qy 330 ValLeuValAspHisAlaIlePheArgAlaPheAsnLysGlyAlaLeuGlyIleLeuLys 349
Db 1492 GCCTACGTCACCAACCTATGATGCGGTTCGAGCTGGGTGCGCGCGCACTTCAAG 1551
Qy 350 ValGluGluGluGluAsnHisGluIleTyrSerHisLysGlnThrAspAlaValTyrLeu 369
Db 1552 GTGACCGCGCAATGGAACGATGATCTG-----ATGACATCGGTGTCTCAAGCG 1599
Qy 370 ProGluGlyAlaProGlnAlaIleAspThrGlnGluAlaProLysThrProAlaProAla 389
Db 1600 GCGTCGATGTACCGCGGTACAGCGATGATGTCGGGCGCGGTACAGCCACCCAGCC 1659
Qy 390 AsnLeuGlnGluGlnIleLysAlaGly-----LysAlaThrTyrAsp 403
Db 1660 GCGAA-CAMAGGACGCGCTTGGCGGCGCTTCCCGGAGACCTTCATGCCACACCA 1718
Qy 404 SerAsnCysAlaIaLysHisGlnProAspGlyLysGlyValProAsnAlaPheProPro 423
Db 1719 GAATACCTGC-----GTTTCCGCTCT 1739
Qy 424 LeuAlaAsnSerAsp-----TyrLeuAsnAlaAspHisAlaArgAlaAlaSerIleVal 441
Db 1740 CTCGCTGACGACACCGGTGCTTTCGCGGCAATCGCGCGTCTTC----- 1790
Qy 442 AlaAsnGlyLeuSerGlyLysIleThrValAsnGlyAsnGlnTyrGluSerValMetPro 461
Db 1791 -----GCGGAAGGTGGGCTGTTCAGACACTCAAGACGCGCATGATCCGTCATCAG 1844
Qy 462 AlaIleAlaLeuSerAspGlnGlnIleAlaAsnVal 473
Db 1845 GCCCGCGGTACCGTGAAGATCGCGCCGACACAGATG 1880

```

```

RESULT 17
LOCUS AE007256/c 10587 bp DNA linear BCT 15-AUG-2001
DEFINITION Sinorhizobium meliloti plasmid pSyma section 62 of 121 of the
complete plasmid sequence.
ACCESSION AE007256 AE006469
VERSION AE007256.1 GI:14523793
KEYWORDS
SOURCE Sinorhizobium meliloti (Rhizobium meliloti)
ORGANISM Sinorhizobium meliloti
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
REFERENCE
1 (bases 1 to 10587)

```

```

AUTHORS Barnett,M.J., Fisher,R.F., Jones,T., Komp,C., Abola,A.P.,
Barlcy-Hubler,F., Bowser,U., Capela,D., Galibert,F., Gouzy,J.,
Gurjal,M., Hong,A., Huizar,L., Hyman,R.W., Kahn,D., Kahn,M.L.,
Kaiman,S., Keating,D.H., Palm,C., Peck,M.C., Surzycki,R.,
Wells,D.H., Yeh,K.-C., Davis,R.W., Federspiel,N.A. and Long,S.R.
TITLE Nucleotide sequence and predicted functions of the entire
Sinorhizobium meliloti pSyma megaplasmid
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 98 (17), 9883-9888 (2001)
MEDLINE 21396509
PUBMED 11481432
REFERENCE
2 (bases 1 to 10587)
Barnett,M.J., Fisher,R.F., Jones,T., Komp,C., Abola,A.P.,
Barlcy-Hubler,F., Bowser,U., Capela,D., Galibert,F., Gouzy,J.,
Gurjal,M., Hong,A., Huizar,L., Hyman,R.W., Kahn,D., Kahn,M.L.,
Kaiman,S., Keating,D.H., Palm,C., Peck,M.C., Surzycki,R.,
Wells,D.H., Yeh,K.-C., Davis,R.W., Federspiel,N.A. and Long,S.R.
Direct Submission
Submitted (29-MAR-2001) Biological Sciences, Stanford University,
371 Serra Mall, Stanford, CA 94305, USA
location/Qualifiers
1. 10587
/organism="Sinorhizobium meliloti"
/mol_type="genomic DNA"
/strain="1021"
/db_xref="taxon:382"
/plasmid="pSyma"
/complement(161..604)
/gene="azul"
/note="synonym: Sma1243"
/complement(161..604)
/gene="azu1"
/function="Small molecule metabolism"
/note="glimer prediction; pseudozurin, also called blue
copper protein, a soluble electron transfer protein
associated with periplasmic nitrite reductase;
Cu-binding/plastocyanin domain; Cu-binding/plastocyanin
domain"
/codon_start=1
/transl_table=1
/product="Azul pseudozurin (blue copper protein)"
/protein_id="AAK65336.1"
/db_xref="GI:14523794"
/transl_table="MRIAAGMAVAAVLAFTGSFADEVRMLKSGEAVFEPA
FVKVNPDSVTFVPTDKSHVETIKDVI PGASAFSKMETVYVTFDVPVGVKCT
PHVGMGVAAVVDAPANVEKVAVALPKKARERLDAALVALAQ"
/complement(715..1362)
/gene="Sma1245"
/complement(715..1362)
/gene="Sma1245"
/function="Miscellaneous; Not classified regulator"
/note="glimer prediction picked ORF start at codon 61;
ATG picked further upstream based on BLASTX of upstream
region, found homology to Pseudomonas"
/codon_start=1
/transl_table=1
/product="putative crp/fnr-like transcriptional regulator"
/protein_id="AAK65337.1"
/db_xref="GI:14523795"
/transl_table="MSDADIDRMIAHAATARRVPGDAVFEGGQRTSFFLLHGRKLV
TOVVEDQOQIIVRVHPGDLFGFKAALQSDYPTATAATSLASPTDLPQVFG
NPHLAATVMTQIGRLREAHTRIRKMTQEVERVAIVAVLISQAKOEKGGVRIDF
PIGRDIAEMTGTTLHTVSRLISLMEKGVVEGGRQLIICDLSGLAALADGARD"
/complement(1987..2901)
/gene="nirV"
/note="synonym: Sma1247"
/complement(1987..2901)
/gene="nirV"
/function="Small molecule metabolism"
/note="glimer prediction; global length homology to NirV
precursor of Pseudomonas, catalyzing first step in
denitrification"
/codon_start=1
/transl_table=1

```

```
/product="NitrV periplasmic nitrate reductase"
/protein_id="AAK5338.1"
/db_xref="GI:14523796"
/translation="MAALSKSRVAPVALATPSALITLACILAEITGLSGAPSGSA
VEBPVTVAPRFRVAVGEFKGVAVDGEVTHMSAPLTIKVTYVADYARCV
AEDACLPAEPEHVPDPAAPATGVSPDQAYAMLSRTGALWVLTDLADYAG
SRFPDADLGEDDASNPALRWLADYRETRAKASRESFORIGHGESETGLSDFGN
WEMTTTCVARTLDRGKVISDASSCGYIATGRRAALSFPVNPKGCGAVGAPP
DNVGRVLVDGWRVAPLLRTLRKGLDV"
/complement(3000..4130)
/gene="nitrK"
/feature="synonym: Sma1250"
/complement(3000..4130)
/gene="nitrK"
/EC_number="1.7.99.3"
/function="Small molecule metabolism; Energy metabolism,
carbon; anaerobic respiration"
/feature="glimmer prediction; Global full length homology of
NitrK-like Cu-containing nitrite reductase
(disimilatory)."
/codon_start=1
/translation="conserved hypothetical protein"
/product="putative NitrK Cu-nitrite reductase"
/protein_id="AAK5339.1"
/db_xref="GI:14523797"
/translation="MSEOFQMTRRSMLAGATVPLIGAVSAHEEAVAKTAMIN
VASLPKRVVDLYKPPFAHDTQKAGGKVEFTLIEKVIDEOTELHMTFNG
SVPGLMVDHODVYELTINDPTTLQHNIDFHSATGALGALTYVNPBDTTLRF
KASKAGVYVHCAPPGVPMHVTSGMNGALNVLREGITDGKNSITTDKYVGEQD
FYPRDANGKEKKEVSEAVADTLEVRITLPSHIVNGAVALTGSALKAAYGEK
VLIVASQANRDTRPHLIGHDGVYVATKFRNAPVDDETFIPGCTAGAAFYTFEOP
GIYAVNHLIEAFELGAAHFAVGTGWNMDLMTSVRAPSGT"
/4376..5587
/gene="Sma1252"
/4376..5587
/function="Miscellaneous; Hypothetical/Global homology"
/feature="glimmer prediction; 30% global identity to ORF 396
in the denitrification gene cluster of Pseudomonas
stutzeri; 46% global identity to NitrS which is found
divergently transcribed from NitrR, a transcription factor
that regulates denitrification in R. sphaeroides"
/codon_start=1
/translation="conserved hypothetical protein"
/product="conserved hypothetical protein"
/protein_id="AAK5340.1"
/db_xref="GI:14523798"
/translation="MSGKRVAKREGYVRFGRFATGVLFSYGRFPFLGAAWAVAM
TLWIALVGHLEVAGSYGAHMAHMEMLFGFAPVLAGFLTAVPNMTGRPLPVSQML
AGLFTLMLAGRAALSPVIGIPPAALIDGLPLALLICAREVJAGKMDLKVIGG
LLATSLANACFPFAVVTGDVHIAWRLGISAVVALVTIIGRIILPSFTRNMLNRAQRT
EPFVYVNFRTYVAILAGTALGATLADHPVTPAPPAALIHTRVLRARRNGRTWP
EMLVILHVAIFVPLGFATGIGALGVSEELSVHVAIVLGVAILAMWLAIVTRARSG
TGYPLTASRLTASAAVVLALRLPLAEMLELAPLTVAVSGSAMILLAFCLIEGH
PILVKKRAVQ"
/complement(5643..6278)
/gene="Sma1254"
/complement(5643..6278)
/gene="Sma1254"
/function="Miscellaneous; Unknown"
/feature="glimmer prediction"
/codon_start=1
/translation="Hypothetical Protein"
/product="Hypothetical Protein"
/protein_id="AAK5341.1"
/db_xref="GI:14523799"
/translation="MSGCFPSKTNASAMOLFHHGSSSRATLFRSLKSPPTFGRRK
LSLVRSRCSTASTISAGSRSCDOTNIDVAGVLTGSGAPVPOHSGNATKAYAG
LYPFDILKPSLAKSWPIPSKCSGLINRPAIPSPSTTRCCGRLSAPRVLP
LTNPILASCAVEMCLPCRQSSLSRRTDSAPKIAFSRYRIVHSAKKAAY"
/5701..5970
/gene="Sma1253"
/5701..5970
```

```
/gene="Sma1253"
/function="Miscellaneous; Unknown"
/feature="glimmer prediction"
/codon_start=1
/translation="Hypothetical protein"
/protein_id="AAK5342.1"
/db_xref="GI:14523800"
/translation="MGASNVRLRLRELCIQRHISATOLAERGLVNSTNALNSLR
OHLVVEBGTGALFRPNSRHPGLDSTIGLFGABSGFKISRG"
/complement(6650..6961)
/gene="Sma1255"
/complement(6650..6961)
/gene="Sma1255"
/function="Miscellaneous; Hypothetical/Global homology"
/feature="glimmer prediction; global similarity to putative
dUDP-4-Keto-L-rhamnose reductase from S. mutans (GI:
7514804)"
/codon_start=1
/translation="conserved hypothetical protein"
/protein_id="AAK5343.1"
/db_xref="GI:14523801"
/translation="MIRAEAMSEPENATVCDISIDALRMIIDPELGRNIVDGLIYDV
SVEDGIAHVMTTITGCPASBYLKEAVRNCWYVGVGEYAEVRLTYEPMTDMMMA
G"
/complement(6948..7478)
/gene="Sma1256"
/complement(6948..7478)
/gene="Sma1256"
/translation="MITETQNPAREDFVRLPSGSGEPFOAIMEAVNGLRPGOALRL
LAPRPOPILFKVMGRGSHAOBIOGGDMEVLPKPAAGAPVVSADADNAPMPD
VENIDILBDDPPEPMYRIILAAVERLQGEVLFALLSBEPIFLPELSKRGQMANFD
ETRTFRIFVRVGDG"
/complement(7482..8840)
/gene="Sma1259"
/gene="Sma1259"
Alignment Scores:
Pred. No.: 1.02e-20 Length: 10587
Score: 456.00 Matches: 133
Percent Similarity: 46.81% Conservative: 65
Best Local Similarity: 31.44% Mismatches: 154
Query Match: 17.36% Indels: 73
DB: 1 Gaps: 16
US-10-088-045-2 (1-502) x AB007256 (1-10587)
QY 60 ProValIIleAPAlaIleValThHISAlaProGU----- 71
DB 4061 CCGCTGATCGCGCGCTGTCTCTCGCCATCGCGAAGAGCGCGTCCGCAAGACCGCCACATC 4002
QY 72 -----ValProProProValAspArgAsp 79
DB 4001 AACGTGCATGCTGCCCCGTCAAGTGCAGCTGGAACCGCCCTTCGTCCACGCC 3942
QY 80 HisPro-----AlaIleValIleValIleValMetGUThValGU--- 92
DB 3941 CATACGAGAGGCGGAGAGTGGCCCAAGGCTCTGACGTTACGCTGACGATGAGAG 3882
QY 93 ---LysValMetCArgLeuAlaAspGlyValGUtyrGlnPheTyrPThrPheGlyGlyGln 111
DB 3881 AAGAAATCGATCGATCGACGAGCAGGCTTACGACGACGACCTTCAACGCTTCG 3822
QY 112 ValProGlyIleMetIleArgValArgGUtyrAspThrIleGlyValGlnPheSerAsn 131
DB 3821 GTCCCCGGTCCGCTGATGATGCTGCACAGGACGATATGTCGAGCTGACGCTCATCAAT 3762
```



mac\_peptide 764..1804  
 /gene="nirU"  
 /product="nitrite reductase"  
 /function="nitrite reduction"  
 /note="Cu-NiR; putative"  
 BASE COUNT 439 a 501 c 554 g 392 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 2, 286-21 Length: 1886  
 Score: 453.00 Matches: 118  
 Percent Similarity: 51.588 Conservative: 45  
 Best Local Similarity: 37.348 Mismatches: 119  
 Query Match: 17,254 Indels: 34  
 DB: 1 Gaps: 10

US-10-088-045-2 (1-502) x PSENIRAX (1-1886)

QY 72 ValProProProValAspArgAspHis-----ProAlaIysVal 84  
 DB 842 GTCAAGCCACCTTGTGACACCCACAGAGCTGAGGGCGAGCCCAAGTTGTC 901  
 QY 85 ValValIysMetGluThrValGluIysValMetArgLeuAlaAsp---GlyValGluTyr 103  
 DB 902 GAGTTCAGATGACCATCCAGAGAAAGAGATCGTTGTGACACAGAGGGGACCGAGTC 961  
 QY 104 GlnPheTrpThrPheGlyGlyGlnValProGlyGlnMetIleArgValArgGluIysAsp 123  
 DB 962 CATGGCATGACCTTCACAGCATCTGTGCGGGCCGATATGATGTTTACACAGATGAT 1021  
 QY 124 ThrIleGluValGlnPheSerAsnHisProAspSerIysMetProHisAsnValAspHe 143  
 DB 1022 TATGTGCAACTGACCTCGTCAATCCCGATACAAAGAAATTCAGACAAATTCGACTTC 1081  
 QY 144 HisAlaIaThrGlyProGlyGlyGlnValGlnValAspPheThrAlaProGlyHisThr 163  
 DB 1082 CATTCGGCAGCGGGTGGCTGGGTGGAGCGCTGACCGTGCATCCGCTACACAG 1141  
 QY 164 SerThrPheSerPheIysValLeuGlnProGlyIleTyrValTyrHisCysAlaValAla 183  
 DB 1142 GCGGTCGTCGCTTCAAGGCCACCAAGCGGTTGTTGTCTATCAGTGT-----GCC 1195  
 QY 184 ProValGlyMet-----HisIleAlaAsnGlyMetTyrGlyIleLeuValGlu 200  
 DB 1196 CCGCGCGCATGTGCGCGCATGTGATGATGCGGCATGATGATGATGATGATGATG 1255  
 QY 201 ProIysGluIleuProIys-----ValAspIysGluTyrTyr 213  
 DB 1256 CCGCGTACCGGTCTCAAGATCAAGAGGCCACGAGCTTTTACGACAGGTTACTAT 1315  
 QY 214 ValMetGlnGlyAspPheTyrThrIys-----GlyIysTyrGly 226  
 DB 1316 GTCCGCGCAGAGCATTTTATGTCCCGAAGATGAGAACGCGCAAGTTCAGAAATACGA 1375  
 QY 227 GlnGlnGlyLeuGlnProPheAspMetGluValAlaIleArgGlu---AspAlaGluTyr 245  
 DB 1376 AGCGCGCGGAGCGCTATCCGATGTGTGTAAGCATGAAACACTGACCCCGACGAT 1435  
 QY 246 ValValPheAsnGlySerValGlyAlaLeuThrGlyValAsnAlaLeuValAlaIysVal 265  
 DB 1436 GTTCGTTTCACAGCGCGGTGCGCGCTGACGCGCGCATGACGCTTTCAGGCAAAAGTG 1495  
 QY 266 GlyGluThrValArgLeuPheValGlyAsnGlyIleProAsnLeuThrSerSerPheHis 285  
 DB 1496 GCGCAT-----CGTCTCTGATTTCTTTCGCAAGCCACAGGATACGCGCCGAC 1549  
 QY 286 ValIleGlyGluIlePheAspIysValHisPheGluGlyGlyValGlyValAsnHis--- 304  
 DB 1550 CTGATCGGGGCGCATGCGCATATATGC---TGGCTACCGCGCAAGTTGCCAACCGCGCG 1606  
 QY 305 -----AsnIleGlnThrThrLeuIleProAlaGlyGlyAlaIleThrGluPheIys 322  
 DB 1607 GAACCTGATCAGAAACCTGTTTCTTCCCGAGGTGTGCGGGCGGCTTACTACAG 1666

QY 323 ValAspValProGlyAspTyrValLeuValAspHisAlaIlePheAspAlaPheAsnIys 342  
 DB 1667 TTCACGACAGCCCGGATCTCTATCCGATGTAAACCAATCTGATGAGCGCTTGAACCTC 1726  
 QY 343 GlyAlaLeuGlyIleLeuIysValGluIleGluIleGluIleGluIle 358  
 DB 1727 GCGCGCGCGCGCATCTCAAGGTGACGCGGCGCATGTGAACGACGACCTG 1774

RESULT 19  
 AF040987  
 LOCUS  
 DEFINITION  
 (nirU), and Nirr (nmr) genes, complete cds.  
 ACCESSION  
 AF040987  
 VERSION  
 AF040987.1 GI:2773327  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Pseudomonas sp. G-179  
 Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 Rhizobiaceae.  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

1 (bases 1 to 4029)  
 Ye, R.W. and Laura, B.  
 Characterization of the DNA region responsible for nitrite  
 reduction in Pseudomonas sp. G-179 containing a Cu-type nitrite  
 reductase  
 Unpublished  
 2 (bases 1 to 4029)  
 Ye, R.W. and Laura, B.  
 Direct Submission  
 Submitted (01-JAN-1998) Environmental Biotechnology, E328/148B,  
 Dupont Central Research & Development, Route 141 and Henry Road,  
 Experimental Station, Wilmington, DE 19880-0328, USA

location/Qualifiers  
 1..4029  
 /organism="Pseudomonas sp. G-179"  
 /mol\_type="genomic DNA"  
 /strain="G-179"  
 /db\_xref="taxon:71238"  
 582..1721  
 /gene="nirU"  
 /note="Cu-type nitrite reductase structural gene"  
 582..1721  
 /gene="nirU"  
 /codon\_start=1  
 /product="nitrite reductase"  
 /protein\_id="AAB96769.1"  
 /db\_xref="GI:2773328"  
 /translation="MSQFRLTRRSMLAGAAVAGALPVVTSVAHAEGGGKXTNSAAT  
 AANATLERVVELVKPVPVHAHQKAGEKVEPMTIOEKIIVDDKTEVHAMT  
 PDGSPGPMIVHODDVELTVNPDPNLEOHNDPSATGALGGGALTVNPDPTV  
 LRFAATAGVFVTHCAPAGVPMHYTSGNMGALIVLPDGLKDKHGLHYDYKYVY  
 EODFYVPKDEGKFKYKESAGEAVPDLVEMKTLTPHVVFNAGVALIGUNALQAV  
 GDRVLIHSQANDRTPHLIGHDVYMAKGPANPELDOETWFIQGAAGAAVYTF  
 QQPGIYAVVHNLIIEAFELGAGHFVTGDMNDIMTAVVSPISG"  
 1803..2699  
 /gene="nirU"  
 /note="similar to cybathionine gamma-synthase"  
 1803..2699  
 /gene="nirU"  
 /codon\_start=1  
 /product="NirY precursor"  
 /protein\_id="AAB96770.1"  
 /db\_xref="GI:2773329"  
 /translation="MTRDAAFRNVLSTAFPALVVGTLGAIQFOSGLPTREGE  
 PVPTIISIPAGLEYSRSEFPNGLVVDAPMTIHSQSLIMKFOVSGSYDCCV  
 GVARPRLSSGRSDAMPATGVSDDAAVYRWISNPGSLWRLPKIETLAPAGSKY  
 DDAIGTIPDSKRNPLRWLADYERETAKARISIPQYGFNGEHEHLADFGGNIW  
 TSTCLRVTLDKRSKPISEASCGIYATGKHGAPLSSFTIREPGGCGVAGPDNW  
 FRLVRDPRWYTPPLFAISRSM"  
 2840..3511  
 gene



CDS

/gene="nmr"  
2840..3511  
/gene="nmr"  
/note="Similar to Nmr from Rhodobacter sphaeroides and other FNR-like regulators"  
/codon\_start=1  
/transl\_table=11  
/product="Nmr"  
/protein\_id="AB96771.1"  
/db\_xref="GI:2773330"

BASE COUNT 864 a 1110 c 1203 g 852 t

ORIGIN

Alignment Scores:

Pred. No.: 5.4e-21 Length: 4029  
Score: 453.00 Matches: 143  
Percent Similarity: 45.93% Conservative: 60  
Best Local Similarity: 32.35% Mismatches: 176  
Query Match: 17.25% Indels: 64  
Gaps: 15

US-10-088-045-2 (1-502) x AF040987 (1-4029)

Oy 72 ValPropProValAspArgAspHis-----ProAlaIysVal 84  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 756 GTCAAGCCACCTTTGTGCACGCCACCAGAAAGCTGAGGGCAGCCCAAGTTC 815  
85 ValValIysMetGluThrValGluLysValMetArgLeuAlaAsp---GlyValGluTy 103  
816 GAGTTCAAAGATGACCATTCACGAAAAAAGAATCGTTTCGACGACAAGGGACCGAGGTC 875  
Oy 104 GluHetrPthrPhneGlyGlyValValProGlyIleMetIleArgValArgGlyIAsp 123  
876 CATCGCATGACGTTTGACGAGATCTGTGCCGGCGCGAGATGAATCGTTCCACAGAGAT 935  
Oy 124 ThrIleGluValGlnPheSerAsnHisProAspSerIysMetProHisAsnValAspHe 143  
936 TATGTCCAACTGACCCCTGTCATAATCCGATACAAACGATTGCACACACATATGACTTC 995  
Oy 144 HisAlaAlaThrGlyProGlyGlyValValAcGluIleAspSerPheThrAlaProGlyHisThr 163  
996 CATTCCGGGACGGGCGCGCTGGTGTGTGACACCGCGATCGATCCCGGTGACACG 1055  
Oy 164 SerThrPheSerPheLysAlaLeuGlnProGlyLeuTyValTYRHisCYsAlaValAla 183  
1056 GC GG TG CG CT TT CA AG GCC A CC A G CG G G GT GT TT TT G T CA T C G T -----GCC 1109  
Oy 184 ProValGlyMet-----HisIleAlaAsnGlyMetTyrglyLeuIleLeuValGlu 200  
1110 CCCGCGGCATATGTCGCCGTGCATGTACTTCGCGGACGATGAATGGCGCATATGTCCTG 1169  
Oy 201 ProIysGluGlyLeuProlys-----ValAspLysGluTyTy 213  
1170 CCGCGTACGGTCTCAAGATCACAGAGCCACGACGCTGTTTACGACAGGTCTACTAT 1229  
Oy 214 ValMetGlnIAspPheTyThrLys-----GlyLysTyrgly 226  
1230 GTCCGGCAGCAGCACTTTATGTATCCCGAAGATGAGACGCGCAAGTTCCAAAGAAATACGA 1289  
Oy 227 GluGlnGlyLeuGlnProPheAspMetCiluYsaIleArgGlu--AspAlaGluTy 245  
1290 AGCGCGCGCAGGCGCTATCCGATGTCTGTGAAGACATGAAGACACTGACCCGACGAT 1349  
Oy 246 ValValPheAsnGlySerValIGLYAlaLeuThrGlyGluAsnAlaLeuValAlaIysVal 265  
1350 GTCCGTTTCMAACGGGGCGGTGGCGCCCTTCAGCGGCATTAACGCTTCAGGCAAGGTC 1409  
Oy 266 GlyIuThrValArgLeuPheValGlyAsnGlyGlyProAsnLeuThrSerSerPheHis 285

Db	1410	GGCGAT-----CGTGTCCGATTTCTTATCCGACGCCAACGAGATACGCGCCGAC	1463
Qy	286	ValIleGIyulIePheAsPlyVaLIaPheGIyulYyGIyulAaSnHs---	304
Db	1464	CTGATCGGGGGGCGATGCGCATTAATGTC---TGCGCTACCGGCAAGTTCCGCAACCCCGC	1520
Qy	305	-----AsnIleGlnThrLleuIleProAlaGIyulAlaIaIeThrGluPheLys	322
Db	1521	GAACCTGCATCAGGAACCTGGTTCATTCCCGAGAGTCTCTCCGGGGCGGCTTACTACCG	1580
Qy	323	ValAspAlProGIyAspTyrValIleValAspHisAlaIePheArgAlaPheAsnLys	342
Db	1581	TTCCAGACAGCCCGGATATCTATGCGTATGTAAACACAACTGATCGAGGGCTTGAATC	1640
Qy	343	GIyAlaLeuGIyIleLeuLysValGIyulGIyulAaSnHsGIyulYySerHisLys	362
Db	1641	GGCGCGCGCGCCACCTTCAGAGTGACGGGCGACTGAACGACGACTGATGACA-----	1694
Qy	363	GlnThrAspAlaValTyrLeuPro-----GluYulAlaProGln	375
Db	1695	-----GCCGTGGTTTCGCCGACCTCGGAGTTGACGATGTCCGCCCGCGCGGATCA	1745
Qy	376	AlaIleAspThrGlnGluAlaProLysThrProAlaProAlaSnLeuGlnGluHle	395
Db	1746	GGCGCGCGGCGCACATCTTATGCCGCGGGGAGCGCGAAGATCGAGAGATCCCATG	1805
Qy	396	LysAlaGIyLysAlaThrTyrAspSerAsnCySAla-----	407
Db	1806	ACCTTTCGCGACGCGTCTTCTT---TCCCAATGTCTTTTTCGATTCCTTTCCGTTCCG	1864
Qy	408	AlaCySHisGlnProAspGIyulYyGIyulProAsnAlaPheProPoleuAlaAsnSer	427
Db	1865	GCTTGTGTGCGGCTTACCGGCG-----AATCGGCTTCCAGTCCGGACTTTTGCC	1915
Qy	428	AspTyrLeu-----AsnAlaAspHis-AlaArgAlaIaSerIleValAl	442
Db	1916	GACCGCGGAAGGGCGCGTCCGTTCCCGGACCATGATATCCCGCAGGTCTTCTGGA	1975
Qy	442	AsnGIyLeuSerGIyLysIleThrValAsnGIyAsnGIyLysSerValMetProAl	462
Db	1976	ATATCCGAGAAAGTGTAGTTCCTTCGCAACGCTCAACCGTGCAGCCGCCGATGACCA	2035
Qy	462	allle 463	
Db	2036	AATT 2039	
RESULT 20			
AE008247			
LOCUS	10484	bp	DNA
DEFINITION	Agrobacterium tumefaciens str. C58		linear BCT 18-DEC-2001
ACCESSION	AE008247		187 of the complete sequence.
VERSION	AE008247.1		GI:15158849
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
COMMENT			
FEATURES			

```
source
1. .10484
/organism="Agrobacterium tumefaciens str. C58 (Cereon)"
/mol_type="genomic DNA"
/strain="C58"
/db_xref="taxon:181661"
67. .1965
/gene="AGR_L_961"
67. .1965
/gene="AGR_L_961"
/notes="(AF083948) TrkA-like protein"
/codon_start=1
/transl_table=1
/product="AGR_L_961p"
/protein_id="AK89054.1"
/db_xref="GI:15158850"
/translacion="MLDTLEETVYGRAMHRLIGETGTLRPPDHAVKLEDMOAVLTICFRGCGEHAQVIAAPARASTHRLRLRQKGLGEKLAQGRDHATLMLEPVLDLPRDRLRNLDLYLMAALMLLEPVADPLKNDLRLKRLKALAEIYTRAFPKKTRYRLCAAMLAIVQTRPLPQAEQVVEQVLFMLAAAGIEGVAPEPASPEPKAPRYLPLPVLPMDTLRETSEARQEDQPAAPAAPOEAAARHIAVEKAENRQTERSPFLINRFEKILAAEMVNVDRPGDSDADGSADELDLTLGRKGRPSARFEPDLDLPEALNHTALTAELTYPEMDYRERYLADHCRVTSQPAKDAETAPEDPTKULIRVRROFVLSRPHELRAOLDGNDLDLAVRARSDIAAGCGSDRIHVASRQANDLYILDVSLSTDANFDRRLVDKEKLLVLAHGLSCGTHSLITTSRRSRVRYEKDFEPMGHTEIRRIALKPGFYTRIGPAIRHSAKLHRRPERRLLLLTDGKNDDVDFGRFAIEDSRRAVSEARRSGSVFQVTVDSKAQSYIPAMFGNGYAIYSRIKLPALP
AIYSLAG"
1965. .2414
/gene="AGR_L_963"
1965. .2414
/gene="AGR_L_963"
/notes="hypothetical protein"
/codon_start=1
/transl_table=1
/product="AGR_L_963p"
/protein_id="AK89055.1"
/db_xref="GI:15158851"
/translacion="MLVYTNLTNLRPVCATMSBNGTADRFQSGAPHRRTSFHPPRWSQMSKQATNIDVRIIPRERHPIFGALGALTEGOSLQITSDHDPRLYLQLATNPGQFGEVLEQGPBEVWRDLIGRLBEQABESSEDMSCGCCGSEH"
2493. .3026
/gene="AGR_L_966"
2493. .3026
/gene="AGR_L_966"
/notes="hypothetical protein"
/codon_start=1
/transl_table=1
/product="AGR_L_966p"
/protein_id="AK89056.1"
/db_xref="GI:15158852"
/translacion="MTWIRELDVRLPLQSGGPPPAIMEAVAGLAPCEALRLYATFRPVPLKMAERGFDOHVEIGDGMDEVLFQPSIETEOISVPALAGNAEBOIWWPDAIVLDSSETBEGCOTSRKSLDLRLPGAVIFVLMQPEPTLYPELLARGHOMAGNDEARTAFRIFIRAGAGAH"
complement (3110. .4384)
/gene="AGR_L_968"
complement (3110. .4384)
/gene="AGR_L_968"
/notes="conserved hypothetical protein PA2662 {imported} - Pseudomonas aeruginosa (strain PA01)"
/codon_start=1
/transl_table=1
/product="AGR_L_968p"
/protein_id="AK89057.1"
/db_xref="GI:15158853"
/translacion="WRLDICODQARSTTMSSEVYPLGSRPKGIPRGLARTGPVIFSYGFRPFLGALMAIVAMVLAISGFLDIGDGGAPQWMAHEMLFGASAVLGGFLTAVPNMTGLPVGSKPLVLMFLALWCAGRLLLVPOVGVVTAADVGLFLPALVIGAREVLAGRKMDLKVGLALASVANIYVHVAIAGDHSOMATRLASVYVTLAVIGRIVPSFTNNMLNRFGRDTPVPVNGSFDAITLITGIVALAWMTIEBSMLAVSTLLAARMAVRLLRMGWTRPEQVIVLHVAVAFIPVGFIAIALADVMQDRSYLHFTVGVICMLAVMVRASLGHTRKLAASRLTIAVVSILACALLRPAEFLPGVMHLYGCSALLMIVGFGFVCEYGPILMRERPLKA"

gene
4354. .5721
/gene="AGR_L_970"
4354. .5721
/gene="AGR_L_970"
/notes="COPPER-CONTAINING NITRITE REDUCTASE PRECURSOR (CU-NIR)"
/codon_start=1
/transl_table=1
/product="AGR_L_970p"
/protein_id="AK89058.1"
/db_xref="GI:15158854"
/translacion="MLGLDTCLDSPSGFPANRQGGATLPPOPSAACQNVAGRRKSLFHLKLYRSATNAVATERNMLCEKRDNRMMETFEHITRNILAGAAFGALAPMIIPSAAGAAEEKKAAPLTSALIALPAKVDLVKPPVHAHTQAKAGCPKIVFTLLTKEOMMLDDKGTVEHMTFNGSVGPMVHODDYVELTLINPTNELQHNIDPHSATGALGGGTLTVNGEXAILRPKATKAGVFVYHCAPEGMVPMHVTSGMGAIMVLPREGLTGCHGKELYVDKYVYGGEDPFIIPREBNRPFKYESAGAMADTLEVMKLTSHLYFNGAVGLTSEHALQAAVGRKVLVHSQANRDTRPHLIGHGDTVATGKFRNPDDIDQETWFIQGTAGAAFYTFEOPGIYAVVNNLIEAFELGAAAHFKVTGEMNNITLMQAVLAPSS1"
5800. .6747
/gene="AGR_L_972"
5800. .6747
/gene="AGR_L_972"
/notes="(AF040987) Nitv precursor"
/codon_start=1
/transl_table=1
/product="AGR_L_972p"
/protein_id="AK89059.1"
/db_xref="GI:15158855"
/translacion="MASPRRQVLMGIPGNSHPENAVSLPLSLATVLTATAIGFQSGIIRSGADTRAMTAPQVYTVPAQFTYRAGEYFDDYAVDGPVLDLTKRONLIMKYOVSSANYARCVAGCGCOQPEPFAAPVGGSIPTAGISODDAHAYAKULSRTGVMRLPSDAELAFAGSLPDDETLGVADSENPAIRMLADYERORSRAAGNPQSGSGFBENEYGLADPGNINEMWTRASCRRITLIGDGEVDTAAACGIPVASGKRRAAMSAFVKAPKSGCAVGPVDNLGFRVLVDTRVYALVEKLRARLIS"
6891. .7598
/gene="AGR_L_974"
6891. .7598
/gene="AGR_L_974"
/notes="(AB031072) FNR-like protein partial sequence"
/codon_start=1
/transl_table=1
/product="AGR_L_974p"
/protein_id="AK89060.1"
/db_xref="GI:15158856"
/translacion="WQYKIDRSVIRSLALFAKADDELGLTHANSRLVPEGEAITEQGSAAHFLILHGRUKVTOVTHDGOIIVRMVPGDLFGFARALORDSDYPTTAAESVILSWOTDLMPQFVONPHLAVSAMQITGRLLEAHTRIREMSTOEVERVHAVALRISOQGRKDLDIRIDFPIRSDIDAMETGTLTHTVSRILSSWESKGLVCGGRQKLLVCNLPGLAQLAEGETSQV"
complement (7669. .8088)
/gene="AGR_L_976"
complement (7669. .8088)
/gene="AGR_L_976"
/notes="CYTOCHROME C2"
/codon_start=1
/transl_table=1
/product="AGR_L_976p"
/protein_id="AK89061.1"
/db_xref="GI:15158857"
/translacion="WRQDKDQDTWRNISIIAAMLAATVPALSEGVDAKGAEPKRCSAHAIGGAKNRVGPOLNGIIGRAAGVPDYTYSAMKKAGGEGLVWTPPELNDFLSAPKKIRGNKALAGISRPEDNLIAIYESTASKPAE"
complement (8134. .8697)
/gene="AGR_L_978"
complement (8134. .8697)
/gene="AGR_L_978"
/notes="hypothetical protein"
/codon_start=1
/transl_table=1
/product="AGR_L_978p"

gene
4354. .5721
/gene="AGR_L_970"
4354. .5721
/gene="AGR_L_970"
/notes="COPPER-CONTAINING NITRITE REDUCTASE PRECURSOR (CU-NIR)"
/codon_start=1
/transl_table=1
/product="AGR_L_970p"
/protein_id="AK89058.1"
/db_xref="GI:15158854"
/translacion="MLGLDTCLDSPSGFPANRQGGATLPPOPSAACQNVAGRRKSLFHLKLYRSATNAVATERNMLCEKRDNRMMETFEHITRNILAGAAFGALAPMIIPSAAGAAEEKKAAPLTSALIALPAKVDLVKPPVHAHTQAKAGCPKIVFTLLTKEOMMLDDKGTVEHMTFNGSVGPMVHODDYVELTLINPTNELQHNIDPHSATGALGGGTLTVNGEXAILRPKATKAGVFVYHCAPEGMVPMHVTSGMGAIMVLPREGLTGCHGKELYVDKYVYGGEDPFIIPREBNRPFKYESAGAMADTLEVMKLTSHLYFNGAVGLTSEHALQAAVGRKVLVHSQANRDTRPHLIGHGDTVATGKFRNPDDIDQETWFIQGTAGAAFYTFEOPGIYAVVNNLIEAFELGAAAHFKVTGEMNNITLMQAVLAPSS1"
5800. .6747
/gene="AGR_L_972"
5800. .6747
/gene="AGR_L_972"
/notes="(AF040987) Nitv precursor"
/codon_start=1
/transl_table=1
/product="AGR_L_972p"
/protein_id="AK89059.1"
/db_xref="GI:15158855"
/translacion="MASPRRQVLMGIPGNSHPENAVSLPLSLATVLTATAIGFQSGIIRSGADTRAMTAPQVYTVPAQFTYRAGEYFDDYAVDGPVLDLTKRONLIMKYOVSSANYARCVAGCGCOQPEPFAAPVGGSIPTAGISODDAHAYAKULSRTGVMRLPSDAELAFAGSLPDDETLGVADSENPAIRMLADYERORSRAAGNPQSGSGFBENEYGLADPGNINEMWTRASCRRITLIGDGEVDTAAACGIPVASGKRRAAMSAFVKAPKSGCAVGPVDNLGFRVLVDTRVYALVEKLRARLIS"
6891. .7598
/gene="AGR_L_974"
6891. .7598
/gene="AGR_L_974"
/notes="(AB031072) FNR-like protein partial sequence"
/codon_start=1
/transl_table=1
/product="AGR_L_974p"
/protein_id="AK89060.1"
/db_xref="GI:15158856"
/translacion="WQYKIDRSVIRSLALFAKADDELGLTHANSRLVPEGEAITEQGSAAHFLILHGRUKVTOVTHDGOIIVRMVPGDLFGFARALORDSDYPTTAAESVILSWOTDLMPQFVONPHLAVSAMQITGRLLEAHTRIREMSTOEVERVHAVALRISOQGRKDLDIRIDFPIRSDIDAMETGTLTHTVSRILSSWESKGLVCGGRQKLLVCNLPGLAQLAEGETSQV"
complement (7669. .8088)
/gene="AGR_L_976"
complement (7669. .8088)
/gene="AGR_L_976"
/notes="CYTOCHROME C2"
/codon_start=1
/transl_table=1
/product="AGR_L_976p"
/protein_id="AK89061.1"
/db_xref="GI:15158857"
/translacion="WRQDKDQDTWRNISIIAAMLAATVPALSEGVDAKGAEPKRCSAHAIGGAKNRVGPOLNGIIGRAAGVPDYTYSAMKKAGGEGLVWTPPELNDFLSAPKKIRGNKALAGISRPEDNLIAIYESTASKPAE"
complement (8134. .8697)
/gene="AGR_L_978"
complement (8134. .8697)
/gene="AGR_L_978"
/notes="hypothetical protein"
/codon_start=1
/transl_table=1
/product="AGR_L_978p"
```

/protein\_id="AAK89062.1"  
 /db\_xref="GI:15158858"  
 /translation="MMNRRELKLIAIATGLPLIGADVLTAAENAKPAGGHHVSE  
 EIRFLDEVAETIIPRTSGAKDAVGAFAVAADYDDEOBALFLSAIPEIKSO  
 ADHKGFELDTAOKKOLLSALDKOABQETPAKPHAFITLVKQTLTLLGFTSTKIGATE  
 VLVYDEIIPGFEDRVPYKKGTPAKGTT"  
 complement (8694..10379)  
 /gene="AGR L 980"  
 /complement (8694..10379)  
 /gene="AGR L 980"  
 /note="probable oxidoreductase chain Cj0415 (imported)"  
 Campylobacter jejuni (strain NCTC 11168)"  
 /codon\_start=1  
 /transl\_table=11  
 /product="AGR L 980p"  
 /protein\_id="AAK89063.1"  
 /db\_xref="GI:15158859"  
 /translation="MADNHYDAIVSGSGISGMAAELOKGLKTLMLBERGNIEVT  
 DYONADKEAMDYPHNRATQEMKAKYPVLSRDYLBELATGMADQETPYVEKRPD  
 WERGYNVGRSLIMGQYRMSTDEPANKDGIADVDPRIYEDVSPWYVERFAGI  
 SGRREGDLIPDEPLPIPLNFVEDVASRLKAKFKGTMLINSCANTOLRPOE  
 RFPCCFRNKCRLGCPFGYVSTOASTLPAVATGNLTLPFSIVKELVDKDKKRG  
 VETDAETLVETVYADIIPLANSSTLNSPVTMNSADYVEGGLSSSGELGNMWDH  
 HRRMGATGVDFEDEFYFKRRPAGYTIPTFRNTGDKKTYLGFQISASRMR  
 ELAEVLIGADYKDTLPDGGWTIGMTAFEMLPYHNRVLDHDKDKMGLPVLMSV  
 EKQNLMDREDMVDAWEMFEAVGIKNVSPRGTYAPGMGIHEMGTARMDPKTSV  
 LKNNQVMDAPNVFTDGAQMTSASCNPSLTLYMALTAARAFAVSEKKNL"

BASE COUNT 2080 a 3290 c 3165 g 1949 t

ORIGIN

Alignment Scores:  
 Pred. No.: 1,61e-20 Length: 10484  
 Score: 453.00 Matches: 148  
 Percent Similarity: 43.43% Conservative: 67  
 Best Local Similarity: 29.90% Mismatches: 168  
 Query Match: 17.25% Indels: 112  
 DB: 1 Gaps: 17

US-10-088-045-2 (1-502) x AE008247 (1-10484)

QY 21 SerGlyCysSerasnGlnAlaAspLySAAlaGlnProLysSerSerThrValaAspAla 40  
 DB 4669 GCGGGCGCGCAAGAAAAGAGCGCGCAAGCCATTGACGAGCGGAGATTGCA 4728  
 QY 41 AlaAlaLysThrAlaAsnAlaAsp-----AsnAlaAlaSerGln 53  
 DB 4729 GCGCTCCCTCGCGCAAGGTTGATCTCGTCAAGCCCTTCTGCGACGCCACACCCAG 4788  
 QY 54 GlnHisGlnGlyGluLeuProValIleAspAlaIleValThrHisAlaProGluValPro 73  
 DB 4789 AAAGCCGAAGCGCGCAAGATCGTCAAGCTTCACGCTCACG 4830  
 QY 74 ProProValaAspArgAspHisProAlaLysValaValaLysMetGluThrValaGluLys 93  
 DB 4831 -----ATCAAG-----GAAACAG 4842  
 QY 94 ValMetArgLeuAlaAsp---GlyValaGluTyrGlnPheThrPheGlyGlnVala 112  
 DB 4843 AAGATATCTTCAGACGCAAGGACCGAGGTTCAAGCTTCACAGCTTCAGAGCTCGGT 4902  
 QY 113 ProGlyGlnMetIleArgValaArgGluGlyAspThrIleGluValaGlnPheSerAsnHis 132  
 DB 4903 CCCGGCCCTTGTATGGTGGTGCATCAGACGACTGAGTCAAGCCCTCATCAACACCC 4962  
 QY 133 ProAspSerLysMetProHisAsnValaAspPheHisAlaAlaThrGlyProGlyGlyGly 152  
 DB 4963 GATACCAAGAGCTGACGCAACATCGATTTCACCTCGGCGCACCGCGCGCTCGGCGCG 5022  
 QY 153 AlaGluAlaSerPheThrAlaProGlyHisThrSerThrPheSerPheLysAlaLeuGln 172  
 DB 5023 GCGGGGCTACCACTCGTCAACCCCGGGAAGGCGGCTCTCGGTTCAGGCGCACCAAG 5082  
 QY 173 ProGlyLeuTyrValTyrHisCysAlaValaIleProValGlyMet-----HisIle 189

DB 5083 GCGGGGCTCTTCGTTTATCACTGC-----GCACCTCCCGGCGATGTCGCGGCAATGTC 5136  
 QY 190 AlaAsnGlyMetCysGlyLeuIleLeuValaGluProLysGlnGlyLeuProLys----- 207  
 DB 5137 ACTCGGGCATGATGATGCGCGATCATGATGCTGCCCGGAGGAGGCTTCAGACGCGCAC 5196  
 QY 208 -----ValaAspLysGluTyrTyrValaMetGlnGlyAspPheTyrThrLys 222  
 DB 5197 GCGAAGAGCTTGTATTATACAAAGCTATTATATGTTGGGAGAGAGACTTCTTACATCCG 5256  
 QY 223 -----GlyLysTyrGlyGlnGlnGlyLeuGlnProPheAspMet 225  
 DB 5257 CGTGACGAGACGCGCAATTTCAGAAATATGAAGCCCGCGTACCCATGCGCGACACG 5316  
 QY 236 GlnLysAlaIleLeuArgLys---AspAlaGluTyrValaValaPheAsnGlySerValaGlyAla 254  
 DB 5317 CTGGAAGTGTATGGCAAGCTGACGACGCAAGCCACATGCTTTCATATGGCGGAGTGTGCC 5376  
 QY 255 LeuThrGlyGlnAsnAlaLeuLysValaValaGlyGluThrValaArgLeuPheValaGly 274  
 DB 5377 CTGACAGGGGAACATCCCTTCAAGCGCGCTGCGTGGTGA-----AAGGCTGATGCTT 5430  
 QY 275 AsnGlyGlyProAsnLeuThrSerSerPheHisValaIleGlyGluIlePheAspLysVala 294  
 DB 5431 CATTCGACGAGCGCAACCGCACACCCCGCGCACCTCATCGCGCGCACGCGCATACGTC 5490  
 QY 295 HisPheGlnGlyGlyLysGlyGlnAsnHis-----AsnIleGlnThrThrLeuIle 311  
 DB 5491 ---TGCGCGACGCGCAAAATTCGCAACCCCGCGATCTCGACGAGAACTGTTTCATC 5547  
 QY 312 ProAlaGlyGlyValaAlaIleThrGlnPheLysValaAspValProGlyAspTyrValaLeu 331  
 DB 5548 CCCGGGGGACAGCGCGCGCGCTTCTTACACTTGGAGAGCGCGGCACTTACGCTTAC 5607  
 QY 332 ValaAspHisAlaIlePheArgAlaPheAsnLysGlyValaLeuGlyIleLeuLysValaGln 351  
 DB 5608 GTCACACCAACACTCATCAGAGCTTCGAACTGGCGCGTGCAGCCCATTCAGAGTCAAG 5667  
 QY 352 GlnGluGlnAsnHisIleGluIleTyrSerHisLysGlnThrAspAlaValaTyrLeuProGlu 371  
 DB 5668 GGTGAAGTGAACAACACCTCG-----ATCGAGGCGGTGCTC----- 5703  
 QY 372 GlyAlaProGlnAlaIleAspThrGln---GlnAlaProLysThrProAlaProAlaAs 390  
 DB 5704 ---GCCCATTCAGACATCTGACGCAACGAGGAGCGCGCGCAC----- 5746  
 QY 390 nLeuGlnGlnGlnIleLysAlaGlyLysAlaThrTyrAspSerAsnCysAlaAcysHis 410  
 DB 5747 -----GCGCGGCGCG 5757  
 QY 410 SerLysProAspGlyLysGlyValaProAsnAlaPheProLeuAlaAsnSerAspTyrLe 430  
 DB 5758 CTCCTCCCTTCGCTCTCTCGCGCTGAAAGCGCGCACTTTCGCTTCAC----- 5812  
 QY 430 uAsnAlaAspHisAlaArgAlaAlaSerIleValaAlaAsnGlyLeuSerGlyLysIleThr 450  
 DB 5812 ----- 5812  
 QY 450 rValaAsnGlyAsnGlnTyrGlnSerValaMetProAlaIleAlaLeuSerAspGln-GlnI 470  
 DB 5813 ---GACGACGAGGAGTTTCATGCGCATTCGCGGCAATATGACATTTCCAAATAGCGCTC 5868  
 QY 470 LeuAlaAsnValIleThrTyrThrLeuAsnSerPheGlyAsn 483  
 DB 5869 TCCTCTTTTTCGCGCTCTCGCTTCGACACAGTCTGCGCAAC 5909

RESULT 21  
 AE009367/c 10753 bp DNA linear BCT 20-DEC-2001  
 LOCUS Agrobacterium tumefaciens str. C58 linear chromosome, section 137  
 DEFINITION of 187 of the complete sequence.  
 ACCESSION AE009367 AE008689





```
Db      1095 CTCCTCCCTTCGCTCTCTCGCGCGGAAAGCAGCGCCCAACCTTGGCGTCCAC----- 1041
Qy      430 uAaMaLaaphiLaLaArgLaLaSerLiLeVaLaLaSenGlyLeuSerGlyLySiLeTh 450
Db      1041 ----- 1041
Qy      450 rVaLaSenGlyAaSenGlyrGlyuSerVaLaMeProAlaLeLaLeuSerAaSerGln-Gln 470
Db      1040 ----GAGCAGCAGCAGGTTCTCATGGCGCATTCGCCGCAATACCAATTCCTCCAAATGCCGTC 985
Qy      470 LeLaAaSenValiLeThrTyThrLeuAaSenSerPheGlyAa 483
Db      984 TCGCTCTCTTTCGCGCTCTCGCTTGCACAGTGTGGCAAC 944

RESULT 22
AF083948/c
LOCUS   20761 bp      DNA      linear      BCT 28-APR-1999
DEFINITION
Pseudomonas sp. G-179 FNR-type regulatory protein (nirX), NirX
(nirV), Cu-type dissimilatory nitrite reductase (nirV), NapB
(napB), ferredoxin-like protein (napF), NapD (napD), periplasmic
nitrate reductase large subunit precursor (napA), periplasmic
nitrate reductase small subunit precursor (napB), cytochrome c
binding protein (napC), anaerobic coproporphyrinogen III oxidase
(hemN), trkA-like protein (nirD), putative chaperone (nirG), nitric
oxide reductase large subunit precursor (norC), NorF (norF), cytochrome c
oxidase small subunit precursor (norC), NorF (norF), calcium or iron-binding
protein (tpc), NirV (nirV), and ferrienterobactin-like protein
(lepA) genes, complete cds.
AF083948
AF083948.1 GI:3925388

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Pseudomonas sp. G-179
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
Rhizobiaceae.
REFERENCE
AUTHORS
1 (bases 1 to 20761)
Bedzyk,L., Wang,T. and Ye,R.W.
TITLE
The periplasmic nitrate reductase in Pseudomonas sp. strain G-179
catalyzes the first step of denitrification
JOURNAL
U. Bacteriol. 181 (9), 2802-2806 (1999)
MEDLINE
99235760
10217771
2 (bases 1 to 20761)
Ye,R.W., Bedzyk,L. and Wang,T.
REFERENCE
AUTHORS
Direct Submission
TITLE
Submitted (12-AUG-1998) Life Sciences, Dupont, Experimental Station
E328/148B, Wilmington, DE 19880, USA
JOURNAL
Location/Qualifiers
FEATURES
source
1..20761
/organism="Pseudomonas sp. G-179"
/mol_type="genomic DNA"
/beta_in="G-179"
/db_xref="taxon:71238"
/note="based on 16S rRNA gene sequence, this strain is
similar to Rhizobium"
complement (519..1190)
/gene="nirX"
complement (519..1190)
/gene="nirX"
/note="nirX; putative regulator of nitrite and nitric
oxide reduction"
/codon_start=1
/transl_table=11
/product="FNR-type regulatory protein"
/protein_id="AAC79437.1"
/db_xref="GI:3925389"
/translacion="MPDELDRLQLQASARPAQDAVFEQGGPAPSYLLHGRLLV
TQVTEGGQIIVVHPGDLFGARALQPDYGTATVAVESVTLAMPETLMSFVHE
NPLATYAMQTIERLEAEAVRIRREMSSTEVEERVAATVIRLAKKAGKQKQKDIRIDF
LISRODIAEMTITLTVSRILTSAMQDLGVQGRQLFKVDISGLETTASGEFRKQ
GTASQ"
gene
CDS
complement (1331..2227)
/gene="nirV"
complement (1331..2227)
/gene="nirV"
/function="unknown"
/codon_start=1
/transl_table=11
/product="NirV"
/protein_id="AAC79438.1"
/db_xref="GI:3925390"
/translacion="MTFRDAAFRRNVFLSIAPFPALVNGLTGAIGFQSGLLPFRERG
PVPATISIPAGLEIYESGFFRNGLTVDAPMTKIRHQSLLEIKFQVSGESYQCAAE
GVCAARDLSSGRADMPATGVSPDDAAAYARWLSNRGELIWLPLKLELAFAAGSKYF
DDAIGIDPDGSKNPLRLMADYERETAKAFRIYPOPYGTGEGNEHGLADFGGNIWEM
TSTCLRRTLDKRSKPISEAESCGGIYATCKHGAPLSSFIREFKGGGCSVGAQPDNNG
FRLVRDPRWTPLIFALISGRSM"
complement (2309..3448)
/gene="nirV"
/note="synonym: nirK"
complement (2309..3448)
/gene="nirV"
/codon_start=1
/transl_table=11
/product="Cu-type dissimilatory nitrite reductase"
/protein_id="AAC79439.1"
/db_xref="GI:3925391"
/translacion="MSEOFRLTFRSMLAGAAVAGALAPVYTSVAHAGGGIKTNSAAT
AAIATLELVEKVEIYLPVPHATOKABEPKVEKMTQEKIIVDDKQCEVHMT
FDQSVGPMYIHQDDYVELTLYNPDTNELQNRIDHSATGALGGALTYVNGDITAV
LRKATKAGFVYHCAQAGVNPWHTVSGNKGALIMVLRDGLKDHKHELVLDVYVVG
EDQFVYKDENGEFKRYESAGEAYPDVLEAMKTLTPTHVFNAGAVALTGDNALQAKY
GDRVLILHSQANDRTPLHIGHGIDYVATGKFAANPELDQETWFI PGGAAGAAVYTF
OQGIYAVVNHNIIEAFELGAAGHFVTDGMDNDLMTAVVSPFSG"
3911..4126
/gene="napB"
3911..4126
/gene="napB"
/function="unknown"
/codon_start=1
/transl_table=11
/product="NapB"
/protein_id="AAC79440.1"
/db_xref="GI:3925392"
/translacion="MAEPIATLEQTSFROKRRRELITFVLAFGIWPVVAVGATYTG
FAVMAVOIYVGPGRPHDITPARNSAE"
4132..4632
/gene="napF"
4132..4632
/gene="napF"
/note="NapF; putative electron transporter"
/codon_start=1
/transl_table=11
/product="ferredoxin-like protein"
/protein_id="AAC79441.1"
/db_xref="GI:3925393"
/translacion="MGQVASTRAAFLTGLBNGGHYPGVYDSEVRNCGCGGDCVD
ACPTKILSRDRIPRTIDFQVGTCTGCGSACBPBPPEPATSPAHAVIADNCA
LNPVDOACRDVCPPTAIRFKPRLGSPFPVLDTDCTCGACISICPQOSITMTANT
MEYAA"
4625..4906
/gene="napD"
4625..4906
/gene="napD"
/codon_start=1
/transl_table=11
/product="NapD"
/protein_id="AAC79442.1"
/db_xref="GI:3925394"
/translacion="MBEPQHSSNVVVVRLEAVERMTAALTEIEGEVHAATGGKIV
IYIEGSSFLIETILIRISAMDVIAAHNVFEQTIAEKMSDDGRTHAA"
4881..7385
/gene="napA"
4881..7385
CDS
```

/gene="napa"  
/note="Napa; the periplasmic nitrate reductase region is required for dissimilatory reduction of nitrate; biochemical analysis also indicates this enzyme plays a primary role in the first step of denitrification and there is a lack of membrane-bound nitrate reductase activity in this organism"

/codon\_start=1

/transl\_table=11

/product="periplasmic nitrate reductase large subunit

/protein\_id="AAC79443.1"

/db\_xref="GI:3925395"

/translation="MTAELTRRDVLAQAQAIAASTAGIAMPAAQSVPGVALEIK  
WSKAPRCFCGTVGVGVKENVVATHDMEAEVRKSTCKVGLSTIKMGKRLT  
TPLRKNGVDPKGEFEPVTEEAADIAEKAKTKLKKEKGPALQPSGQMTPEG  
YAATKLMRGPFRSNNIDPNAHCHASATAGPMTFGMDPMGCYDDFEADAFVLMGS  
NNAEMHPIWTRLDRLGHEHVKSVALSTPFRSMDLADPLVEKPGTDLILNVTIA  
NHIQTGRVNOEFIDKNTKFMQATTDIGYLAHEHLEVKATGAAGAAMTPIDEAF  
KVHSELYLEKVAELAGVDKGLAEIADPKVYVSLMTMGAEQAHVGAQOMV  
YVHLLTKISRGNSPFSLTGQPSACGTAREGFAHLPLADMTVTNHEHKAHEL  
KNIPHGIIPEKPGYHAVOODRLHDKLNPVWVYVNNQALANNSEAYIGRPDN  
FIVSDAIPYTVTAMTADILPAAHTEKGAATGAEKRTTHWHQLVNAPGEARSLMO  
LVEFSKRTTDEVPQDILDONPEYKKTLYDVLFGQNVDFPVSEISSDYENREAK  
AFGLYQKGLPEYVAFSGRHGHDLPYDLVHQVRLRPVNNQDETKRKYEGDYP  
VEGEGVYCGNDGRAVILAAPEBPASPDDEFGFVLTGVRLEHMGSGYKMP  
ELYKAPGARCFMNGDDARLGIQNGGVKIOSRGEIISRDIENRNMPIGVIFVP  
WPDASQLINKYLDATDPIKQTDPKKCAVKILPVA"

7355..7846

/gene="napb"

7355..7846

/gene="napb"

/note="NapB"

/codon\_start=1

/transl\_table=11

/product="periplasmic nitrate reductase small subunit

/protein\_id="AAC79444.1"

/db\_xref="GI:3925396"

/translation="KRSQPSRRLSRRLTLPALACLVTVTAALQVTPQISGRSP  
MONTGADLPKMIVDIQTMRNYPDQPIPHSIEGYQSVTNTRMCSCHRELTGGS  
GAPMI SVTHYMRBEGOMLADVSPRYFCTACHVPQADTRPLVDNTEFKMSDELFGPAG  
SQQ"

7843..8547

/gene="napc"

7843..8547

/gene="napc"

/note="NapC; electron transport for nitrate reduction"

/codon\_start=1

/transl\_table=11

/product="cytochrome c binding protein"

/protein\_id="AAC79445.1"

/db\_xref="GI:3925397"

/translation="MIGMKAVILWAMRIATPAGTLGFLTLGGFVGVIWGAEN  
TALVETNTEAFCGCGHEMKTNYVELTVHNSRNGVASCPCDCHPOMDTLARK  
MOASKVEWKGKRTGISTREKFLDKRELAELOHMAWKANDSLKCNCHSITAMDITKO  
TKRAADIRHRYILSGKATCIDCHKGIAMHLEPMWDGVDPWKYPALMNGAASHGLQDT  
DOLAAYLAKESNIID"

9696..11048

/gene="hemn"

9696..11048

/gene="hemn"

/function="anaerobic heme biosynthesis"

/codon\_start=1

/transl\_table=11

/product="anaerobic coproporphyrinogen III oxidase"

/protein\_id="AAC79446.1"

/db\_xref="GI:3925398"

/translation="MKRSILERYGDARLEPRYSYTPARFSSISAGTYSMDLSHDD  
GPISLYVHAIPIPCRSWCVGCHTITLADQPILDLVDLKRKESILVADMPKPEIA

Prod. No.: 3.5e-20 Length: 20761  
Score: 453.00 Matches: 143  
Percent Similarity: 45.93% Conservative: 60  
Best Local Similarity: 32.35% Mismatches: 176  
Query Match: 17.25% Indels: 64  
DB: 1 Gaps: 15

US-10-088-045-2 (1-502) x AF083948 (1-20761)

QY 72 ValProProValAspArgAspHis-----ProAlaValVal 84  
DB 3274 GTCAAGCACCTTTGTGTGACCCCAACCCGAAAGCTGAGGCGGACGCTTTC 3215  
QY 85 ValValValMetGluThrValGluValValMetArgValValAsp---GlyValGluThr 103  
DB 3214 GAGTTCAGATGATACCATCCAGAAAAAGAGATCGTTGTGACAGCAAGGCGACGAGTC 3155  
QY 104 GluMetPheThrPheGlyGlyGluValProGlyGluMetIleArgValArgGluGluAsp 123  
DB 3154 CATGCGATGACCTTCGACGAGATCTGTGCGGCGCGATGATGATGATGATGATGATGAT 3095  
QY 124 ThrIleGluValGluPheSerAsnHisProAspSerIleMetProHisAsnValAspPhe 143  
DB 3094 TATGTGAACTGACCTCGTCAATCCGATACAAAGAAATTCAGCACAATATCGACTTC 3035  
QY 144 HisAlaAlaThrGlyProGlyGlyGlyAlaGluAlaSerPheThrAlaProGlyHisThr 163  
DB 3034 CATTCGGGACGGGTCGCGTGGTGGTGGAGGCGGTGACCGGTGTCATCCGGTGCACAG 2975  
QY 164 SerThrPheSerPheValAlaGluGluPheValValValValValValValValVal 183  
DB 2974 GCGGTGCTGCTTTCAGGCGCACAGGCGGCGGTGTTTGTCTATCATCTGT-----GCC 2921  
QY 184 ProValGlyMet-----HisIleAlaAsnGlyMetTyrGlyIleuIleuValGlu 200  
DB 2920 CCGCGCGGATGTCGCGTGCATGTCCTTGGGATGATGATGATGATGATGATGATGATGAT 2861  
QY 201 ProValGlyGlyLeuProVal-----ValAspValGlyTyrTyr 213  
DB 2860 CCGCGTGAAGCTCTCAAGATACAAAGGCCGACGCTTGTACGACAAAGCTCTATCT 2801  
QY 214 ValMetGluGlyAspPheThrThrHis-----GlyIleTyrGly 226  
DB 2800 GTTCGCGACGACGACCTTTATATGCTCCGAAAGATGACAGACGCAAGTTCAGAAATGACA 2741  
QY 227 GluGluGlyLeuGluPhePheSerMetGluValAlaIleArgGlu---AspAlaGluTyr 245  
DB 2740 AGCGCGCGGAGGCTTATCCGATGTCGTAAGCCATGAACACTGACCCGACGACAT 2681  
QY 246 ValValPheAsnGlySerValGlyAlaLeuThrGlyGluAsnAlaLeuValAlaValVal 265  
DB 2680 GTTCGTTTCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2621  
QY 266 GlyGluThrValArgLeuPheValGlyAlaGluGlyProAsnLeuThrSerSerPheHis 285  
DB 2620 GCGCAT-----CGTGTCTGATCTTCAATTCGCAAGCAAGGATACGCGCGCGGCGG 2567  
QY 286 ValIleGlyGluIlePheAspValHisSerGluGlyGlyValArgGluGluHis--- 304  
DB 2566 CTGATCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2510  
QY 305 -----AsnIleGluThrThrIleuIlePheProAlaGlyGlyValAlaIleThrGluPheHis 322  
DB 2509 GAATCGATCAGAAACCTGCTTCTTCCCGAGGTCGTCGCGGCGGCGGCGGCGGCGGCGG 2450  
QY 323 ValAspValProGlyAspTyrValLeuValAspHisAlaIlePheArgAlaPheAsnVal 342  
DB 2449 TTCACGACGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2390  
QY 343 GlyAlaLeuGlyIleLeuValValGluGluGluGluGluGluGluGluGluGluGluGlu 362  
DB 2389 GCG 2336

QY 363 GlnThrAspAlaValTyrLeuPro-----GluGlyAlaProGln 375  
 Db 2335 -----GCCGTGGTTTCGCCGACCTCGGGTTGACGGTGTGCGCCGCCGAGTCA 2285  
 QY 376 AlaIleAspThrGlnGluAlaProLysThrProAlaProAlaLeuGlnGluIle 395  
 Db 2284 GGGCGCCGGCCACATCTTATGCGCGCGGGGAGCGGAGAGATCGAGATCCCATG 2225  
 QY 396 LysAlaGlyLysAlaThrTyrAspSerAsnGlyAla----- 407  
 Db 2224 ACCTTTCCGACGCTGCTTCTT-TCCGATGTGTCTTTTTCGATGCTTTCCGTTCCG 2166  
 QY 408 AlaCysHisGlnProAspGlyLysGlyValProAsnAlaPheProProLeuAlaAsnSer 427  
 Db 2165 GCTTGTGTGCGGCTGACCGCGCC-----AATCGGCTTCCAGTCCGACCTTTTGCC 2115  
 QY 428 AspTyrLeu-----AsnAlaAspHis-AlaArgAlaAlaSerIleValAl 442  
 Db 2114 GACGCGCGAAGGGCGGCTCGGCTTCCGCGACATCATATCCCGGCTCTTCGGA 2055  
 QY 442 AsnGlyLeuSerGlyLysIleThrValAsnGlyAsnGlnTyrGluSerValMetProAl 462  
 Db 2054 ATATCGAAGAAAGTGTAGATTCTTCCGACAGCTCTCACCGTCGACCGCGATGACCA 1995  
 QY 462 aile 463  
 Db 1994 AATT 1991  
 RESULT 23  
 AF051831 1709 bp DNA linear BCT 05-MAR-1999  
 LOCUS Alcaligenes xylosoxidans dissimilatory copper-containing nitrite  
 DEFINITION reductase (nir) gene, complete cds.  
 ACCESSION AF051831  
 VERSION AF051831.1 GI:2967826  
 KEYWORDS  
 SOURCE Achromobacter xylosoxidans  
 ORGANISM Achromobacter xylosoxidans  
 Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
 Alcaligenaceae; Achromobacter.  
 REFERENCE  
 AUTHORS Suzuki, E., Horikoshi, N. and Kohzuma, T.  
 TITLE Cloning, sequencing, and transcriptional studies of the gene  
 encoding copper-containing nitrite reductase from Alcaligenes  
 xylosoxidans NCIMB 11015  
 JOURNAL Biochem. Biophys. Res. Commun. 255 (2), 427-431 (1999)  
 MEDLINE 99160880  
 PUBMED 10049725  
 REFERENCE  
 AUTHORS Suzuki, E., Horikoshi, N. and Kohzuma, T.  
 TITLE Direct Submission  
 JOURNAL Submitted (03-MAR-1998) Environmental Sciences, Ibaraki University,  
 2-1-1 Bunkyo, Mito, Ibaraki 310-8512, Japan  
 FEATURES  
 source  
 1. 1709  
 /organism="Achromobacter xylosoxidans"  
 /mol\_type="genomic DNA"  
 /strain="NCIMB11015"  
 /db\_xref="taxon:85698"  
 348..>1662  
 /gene="nir"  
 /gene="nir"  
 348..360  
 /gene="nir"  
 /note="fur-box"  
 435..462  
 /gene="nir"  
 /note="nirA-box"  
 488..>1662  
 /gene="nir"  
 /product="dissimilatory copper-containing nitrite  
 reductase"  
 568..573  
 /gene="nir"  
 misc\_feature  
 435..462  
 /gene="nir"  
 mRNA  
 488..>1662  
 /gene="nir"  
 /product="dissimilatory copper-containing nitrite  
 reductase"  
 568..573  
 /gene="nir"  
 RBS  
 568..573  
 /gene="nir"

CDS  
 580..1662  
 /gene="nir"  
 /codon\_start=1  
 /transl\_table=11  
 /product="dissimilatory copper-containing nitrite  
 reductase"  
 /db\_xref="GI:2967827"  
 /protein\_id="AAC05831.1"  
 /translation="NNALRPTLLAALAFMAAGTAAODADKLPHTKYTLVAPQVH  
 PHEQATKSGPKVEFTMTIEKKMVDIDGCTTQAMTFNPSMGPTLVNHEGYVOLT  
 LVNPAFNAPHNVDPRGATGALGALTNVNBESQTLPRKDRSGTFVYHCAPEGVN  
 PMHVVGMSGGLLVLPRLDKDPQGRPLHYDRVYTGEPDLVYPKGSDYKQYATLA  
 ESYGDVQWVRITLTPSHIVNGKVGALTYANALTAAGVETVLLIHSQARDYTPHIG  
 GHGDVWETGKEFANPPQDLIEWFIRGSGAALYFKQPGVAYALNHNLEAFELGA  
 AGHIKVEGKMDLMKQIKAPAPIPR"  
 BASE COUNT 333 a 577 c 560 g 239 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 2,37e-21 Length: 1709  
 Score: 452.00 Matches: 135  
 Percent Similarity: 49.73% Conservative: 49  
 Best Local Similarity: 36.49% Mismatches: 142  
 Query Match: 17.21% Indels: 44  
 DB: 1 Gaps: 16  
 US-10-088-045-2 (1-502) x AF051831 (1-1709)  
 QY 41 AlaIleAspThrGlnGluAlaProLysThrProAlaProAlaLeuGlnGluIle 60  
 Db 631 GCGCGCGGACCGCTCGGCGGCGAGAGCC-----GACACGCTGCC 672  
 QY 61 ValIleAspAlaIleThrHisAlaProGlnValPro-----ProValaParg 78  
 Db 673 CATACCAAGTCAAGCTGTGCTCGCTCCCGCCAGCTCATCTCGACAGACGCGACCAAG 732  
 QY 79 AspHisProAlaLysValValValValValValValValValValValValValVal 98  
 Db 733 TCGCGCGCCAAAGTGTGATTCACCATGACCATGAGAGAAAGATGCGATCGAC 792  
 QY 99 Asp--GlyValGluTyrGlnPheThrPheGlyGlyGlnValProGlyGlnMetIle 117  
 Db 793 GACAAAGGACGACGCTGACGAGCGATGACGTTCAAGGCTCCATGCCCGCCGACGCTG 852  
 QY 118 ArgValArgGluGlyAspThrIleGlnValGlnPheSerAsnHisProAspSerLysMet 137  
 Db 853 GTGTCGACAGGAGCGACCTACGCTGACGCTGACGCTGACGCTGACGCTGACGCTGAC 912  
 QY 138 ProHisAsnValAspPheHisAlaIleThrGlyProGlyGlyValAlaGluAlaSerPhe 157  
 Db 913 CGCACAAAGTGTGATTCACCATGACCATGAGAGAAAGATGCGATCGAC 972  
 QY 158 ThrAlaProGlyHisThrSerThrPheSerPheLysAlaLeuGlnProGlyLysVal 177  
 Db 973 GTCAACCTGCGACGACGCTACGCTGCTTCMAAGCGACGCGACGCTGCTGCTGCTGCT 1032  
 QY 178 TyrHisCysAlaValAlaProValAlaGlyMet-----HisIleAlaAsnGlyMetTyr 194  
 Db 1033 TACCATGCTC-----CGCCCGAAGGACATGTCCTGCGACGCTGCTGCGACATGAC 1086  
 QY 195 GlyLeuIleLeuValGluProLysGluGlyLeuPro-----Lys 207  
 Db 1087 GGCACGCTGATGTGTGCTGCGCGACGCGCTGAAAGATCCGACGAGGCAAGCGCTGAT 1146  
 QY 208 ValAspLysGluTyrTyrValMetGlnGlyAspPheTyr-----ThrLysGly 223  
 Db 1147 TACGACCGCGCTACACATCGCGAGTTCGACTGTATCATCCCAAGGCGCGGACGCG 1206  
 QY 224 LysTyrGlyGluGlnIleLeu-----GlnProPheAspMetLysAlaIleArg 240  
 Db 1207 AAGTACAGACATCAAGCGCTGCGCAAGAGTATGAGACGCTGACGATGATGCGC 1266  
 QY 241 Glu--AspAlaGluTyrValValPheAsnGlySerValGlyValaLeuThrGlyLysAsn 259





```

Db      598 ATCGGCGAGTCGACCTGTATATCCCAAGACAAAGCGGCCACTACAGACTACCG 657
        233 Pheaphectgluylalailearglu-----AspAlaGlu 244
        658 ---GACCTGGGCTCAGTTACAGAGACCCCGCGGTATGCCACCTGACCCCGAGC 714
Qy      245 TyrValValPheAenGlySerValGlyAlaLeuThrglyUaenAlaLeuLysAlaLys 264
        715 CACGTGGTGTTCACAGCGCGGTGTCGCGCGCTGACCGGCGCAACGCGCTGACCTCCAG 774
Qy      265 ValGlyLthrValAlaGluPheValGlyAenGlyProAenLeuThrSerPhe 284
        775 GTCCGGAGAGCCGCG---CTGTTATC---CACGCCAGGCCAACCGGACGCGCTGCC 828
Qy      285 HisValIleGlyGlyIlePheApyLysValHisPheGlyGlyLysGlyLys--- 303
        829 CACCTGATTGGCGGCCACGCGGCTGCTG---TGCACACCGGCGCAAGTTCGCAACCG 885
Db      304 -----HisenIleGlnThrThrLeuIleProAlaGlyGlyAlaAlaIleThrgluPhe 321
        886 CCGCAACGCAACATGAAACCTGGTTATCCCTGGCGTTCGCGGTGCGCGCGCTGAC 945
Qy      322 LysValAlaPylProGlyApyThrValLeuValAspHisAlaIlePheArgAlaPheAsn 341
        946 ACCTTCAAGCAGCGCGGACCTACCTGTAACCTGACCATTAACCTGATGAGCCATGAA 1005
Qy      342 LysGlyAlaLeuGlyIleLeuLysValGlyGlyGlyGlyGlyGlyGlyGly 358
        1006 CTGGGGGCGCTCGCGCCAGATCAAGTGTGAGGGGAGCGGACGATGACCTG 1056

RESULT 25
AP005961 298550 bp DNA linear BCT 28-MAR-2003
LOCUS      Bradyrhizobium japonicum USDA 110 DNA, complete genome, section
DEFINITION
ACCESSION AP005961 BA000040
VERSION    AP005961.1 GI:27355365
KEYWORDS
SOURCE     Bradyrhizobium japonicum USDA 110
           Bradyrhizobium japonicum USDA 110
           Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
           Bradyrhizobiaceae; Bradyrhizobium.
REFERENCE  1
           Kaneko,T., Nakamura,Y., Sato,S., Minamisawa,K., Uchiyumi,T.,
           Kohama,M., Matsunobe,A., Ideasa,K., Iriuchih,M., Kawashima,K.,
           Kohara,M., Matsunobe,M., Shimo,S., Tsurunaka,H., Wada,T., Yamada,M.
           and Tabata,S.
           Complete genomic sequence of nitrogen-fixing symbiotic bacterium
           Bradyrhizobium japonicum USDA110
           DNA Res. 9 (6), 189-197 (2002)
           22484998
           12597275
TITLE      Bradyrhizobium japonicum USDA110 (supplement)
JOURNAL    MEDLINE
PUBMED     22485002
REFERENCE  3 (bases 1 to 298550)
AUTHORS    Kaneko,T.
TITLE      Direct Submission
JOURNAL    Submitted (20-NOV-2002) Takakazu Kaneko, Kazusa DNA Research
           Institute, The First Laboratory for Plant Gene Research, 2-6-7
           Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan
           (E-mail: kaneko@kazusa.or.jp,
           URL: http://www.kazusa.or.jp/rhizobase/
           Tel: 81-438-52-3935 (ex. 2338), Fax: 81-438-52-3934)
           Location/Qualifiers
           1. 298550

gene
/organism="Bradyrhizobium japonicum USDA 110"
/mol_type="genomic DNA"
/strain="USDA110"
/db_xref="taxon:224911"
11..721
/gene="nmrR"
11..721
/gene="nmrR"
/feature="Cp family"
ORF ID: b1r7084
/codon_start=1
/transl_table=11
/product="transcriptional regulatory protein"
/protein_id="BAC52349.1"
/db_xref="GI:27355366"
/transl_table="MakutsuYAHU.PI.PAGYTPPALDELTPARSARYPKNSAIFQ
GADQSFLLHCHVRAKATTPGEQIVRYVAPGETFGIAAIGARPYATMAVD
SVVLTPSWPRLVERFPALANTQVTRLOESRILGEMSTQVEQRALHLLR
LAQSGKLDHGIIDPISRODIAQMTGTLHTVSRILSGMEQGLVSGRORILLR
EPHRIIVLAERSPSGAP"
complement (842..1039)
/gene="b1r7085"
complement (842..1039)
/gene="b1r7085"
/feature="ORF ID: b1r7085"
/hypothetical protein"
/codon_start=1
/transl_table=11
/protein_id="BAC52350.1"
/db_xref="GI:27355367"
/transl_table="MPFGSDLDVDDIMRTAPTHIRVLAFLACVGCPIATFHTVEDA
CREHGIDRDKFLALCDVPA"
complement (1297..2649)
/gene="hemN"
complement (1297..2649)
/gene="hemN"
/feature="ORF ID: b1r7086"
/codon_start=1
/transl_table=11
/product="anaerobic coproporphyrinogen III oxidase"
/protein_id="BAC52351.1"
/db_xref="GI:27355368"
/transl_table="MRADLAVSYGEERLPRYTSYPAPHPSPVIDAGYARWLSEDA
GASASLYLVPFCREMCWYCGCTQIVRRDDILIAAYORTLRSEIALVAETIGRIKVE
HIFGGGTPTLMAPEAFELMAMRQAFVLPASIAVEIDPTLPADVEARLISG
NARSIGOSFDPYQVAGINRISFEGTAAVMDLRBAGIAGINFDLIYGI.PHOTVASC
LDTVRSLLAPDRSVFGYAHVPDPKQIKRMINGALPDGPARHQAICAIYALYKA
GYVQIGLDHFAFDDSMAVAFFERTLRNFGQTTDQGEVLGFGASAIQILPQGVQ
NEVOIGAVYOSIGASRLATAGYGLTDDRLRADIIERTINCEFSADLIGICARHGEF
EAWLKASRLKPLISGVVRLDGDRLAVANDSRFLVRSVAALFADHLDPKOLHSBAY"
2853..3053
/gene="b1r7087"
2853..3053
/gene="b1r7087"
/feature="ORF ID: b1r7087"
/unknown protein"
/codon_start=1
/transl_table=11
/protein_id="BAC52352.1"
/db_xref="GI:27355369"
/transl_table="MRLALELILFSLLDVRGHI PRFDPFPVPASPAPAGIVHTLAAV
IVTLAVLSLAINAAVVAIHL"
3079..3954
/gene="b1r7088"
3079..3954
/gene="b1r7088"
/feature="ORF ID: b1r7088"
/unknown protein"
/codon_start=1
/transl_table=11
/protein_id="BAC52353.1"
/db_xref="GI:27355370"

```

FEATURES

Source



[illegible]

Journal	Appl. Environ. Microbiol.	(1996)	In press
REFERENCE	2 (bases 1 to 1499)		
AUTHORS	Toftanin, A. and Shapleigh, J.P.		
TITLE	Direct Submission		
JOURNAL	Submitted (31-JUL-1996)	Microbiology, Cornell University, Wing Hall, Ithaca, NY 14853, USA	
FEATURES	Location/Qualifiers		
source	1..1499	/organism="Rhizobium euliiae"	
		/mol_type="genomic DNA"	
		/strain="HCONT1"	
		/db_xref="taxon:50338"	
protein_bind	40..52	/note="similar to Fur binding motif in Escherichia coli"	
		/bound_molety="FixK"	
gene	119..1264		
		/gene="nirK"	
RBS	119..122		
		/gene="nirK"	
	131..1264		
		/gene="nirK"	
		/codon_start=1	
		/transl_table=11	
		/product="copper containing nitrite reductase"	
		/protein_id="AAB05880.1"	
		/db_xref="GI:1488172"	
		/translation="MTNTLTQTRRTMTLGAAGALPTILTSGGASPTPVRLSAN EIALPRKRLDLYKPEVHVHTQAEAGPKEVETTLTIEKKLVIDKGTEVAMTFD GAVGGLPIVHVDHYEVTLVNPEPTLNHINIDFHSATGALGGALRVNPGESAVLR FKATKAGFVYHCAPGCMVPMVHTSGNNGAIVMLPRGLDGHCKELVDPKVVYLGSC DPTIPRDEKSEPKFYKDSPEGAIEDPTVAVMKTLPTPHIVFNAGALTEGNALTAVEES RVLIHVSQANRDLTPHLIGHGEIVMTGKFWVPDSDQETWFIPEPTRAAAYTFPEQ PGIYAYVNHMLIEAFELGAAAHFVKTGDMDMDLTMTVRSFGS"	
BASE COUNT	323 a	443 c	428 g
ORIGIN			305 t
Alignment Scores:			
Pred. No.:	1:3e-20	Length:	1499
Score:	440.00	Matches:	135
Percent Similarity:	44.34%	Conservative:	61
Best Local Similarity:	30.54%	Mismatches:	165
Query Match:	16.76%	Indels:	82
	1	Gaps:	15
US-10-088-045-2 (1-502) x RHU65658 (1-1499)			
Qy	1 MetSerlySProthThrLeuIlelySthThrThleuIleCys---	AlaLeuSerAlaLeuMet	19
	149 ATGACCCGGGCGACGAGTCTGACGGCGCTGCTGTCGCCGAGGCTGACGCCCATCTC	208	
Qy	20 LeuSerlyGlySerLeuGlnAlaAspLysAlaAlaGlnProLysSerSerThrValaAsp	39	
	209 ACCTCGGGGGGGGGAATGCCCTCCGCCACCTCTGACGAAGCTCACGCCACCTGAATC	268	
Qy	40 AlaAlaAlaIlySthThrAlaAsnAlaAspAsnAlaAlaSerGlnIuHISGlnGlyLeu	59	
	269 GCTGCCTTCGCCGCGAAGAACTGAC-	295	
Qy	60 ProValIleAspAlaIleValThrHisAlaProGluValProProProValaAspArgAsp	79	
	296 -----	CTCGTAAGCCACCTTTCCTGCACGTA	322
Qy	80 HisPro-----	AlaIlyValIValIlySmetGluThrValaGluLys	93
	323 CATACGAAAAAGCAGAGCGGACCGAAGAGTCTGCAAGTACACCTGACGATGAGAA	382	
Qy	94 ValMetArgLeuAlaAsp-----	GlyValaGluTyGlnPheTrpThrPheGlyGlyGln	111
	383 AAGAGCTCGTATCGATGCGAAGGCGACCGGACGTAAGTCAATGACCTTGAACGATCG	442	
Qy	112 ValProGlyGlnMetIleArgValaArgGlnGlyAspThrIleGluValaGlnPheSerAsn	131	
	443 GTTCTGTCGTCGGCTGTTGTCATCAGAGACGACCTCGAGGTCAACCTGTCAT	502	

```

QY      132 HisProAspSerLysMetProHISAsnValAspPheHisAlaIleThrGlyProGlyGly 151
DB      503 CCGGAGCAAGCAACGCGTGCAGACACATGATTCACATGACGACACCGGTGGCTGGC 562
QY      152 GYAlaGluAlaSerPheThrAlaProGlyHisThrSerThrPheSerPheValAlaLeu 171
DB      563 GGGGGTGCACCTGACGCTGTCACACCGCGGTGAGAGCGCGGTATTCGCTTCAAGCGCAC 622
QY      172 GluProGlyLeuThrValTyrHisCysAlaValAlaProValGlyMet-----His 188
DB      623 AAGCGGGGCTCTTCTGCTATCACTGC-----GCCCTCCCGGATGATGATCTTGGCAC 676
QY      189 IleAlaAsnGlyMetTyrGlyLeuIleValGluProLysGlyLeuProLys--- 207
DB      677 GTAACTCGGCGATGAACGGCGGATGATGCTGCTGGCGGCGAAGGTCTGACAGATGCG 736
QY      208 -----ValAspLysGluTyrTyrValMetGlnGlyAspPheTyr--- 220
DB      737 CATGGCAAGAACTGTCTACCAAGGTCTATTACTCGGTGAGCAGACCTTTATATT 796
QY      221 -----ThyLysGly-----LysTyrGlyGluGlnGlyLeuGlnProPheAsp 234
DB      797 CCGCGGAGCAAAAGGCGGAGTTCAAAAATATGACAGCCCGGTGAGGCTTATGAGGAT 856
QY      235 MetGluLysAlaIleArgGlu---AspAlaGluTyrValValPheAsnGlySerValGly 253
DB      857 ACCGTGGCGCTCATGGCTACTTGACCCGACGATATTCTTCAATGCGCGGTGGCG 916
QY      254 AlaLeuThrGlyGluAsnAlaLeuLysValLysValGlyGluThrValArgLeuPheVal 273
DB      917 GCGCTTACCGGTGAAAATGCTTGACGCGTGGCGGCGAG-----CGGTTCTCATC 970
QY      274 GlyAsnGlyGlyProAsnLeuThrSerSerPheHisValIleGlyGluIlePheAspLys 293
DB      971 GTTCACTCGGACCAACCGTGAACCCGCGCATCTTATC----- 1012
QY      294 ValHisPheGluGlyGlyLysGlyGly----- 302
DB      1013 -----GGCGGTATGGCGAATATGCTGGCGGACCGGCAAGTCTGCAACGTG 1060
QY      303 ---AsnHisAsnIleGlnThrThrLeuIleProAlaGlyAlaIleAlaIleThrGluPhe 321
DB      1061 CCGGACCGGATCAGAGACCTGCTCATTCGCGGCGCCACACGCGGCGGCGCTACTAT 1120
QY      322 LysValAspValProGlyAspTyrValLeuValAspHisAlaIlePheAlaPheAsn 341
DB      1121 ACCTTGAACAGCCCGGTATCTATGCTATGCAACCATATCTGATCGAGGCTTTCGAG 1180
QY      342 LysGlyAlaLeuGlyIleLeuLysValGluGlyGluAsnHisGluIleTyrSerHis 361
DB      1181 CTGGCGCTCGGCTCACTTCAAGGTACCGGTGATGGAACGACATCTTAAGACCAACC 1240
QY      362 LysGlnThrAspAla-----ValTyrLeuProGluGlyAlaProGluAla 376
DB      1241 GTGAGGTGCGGATGAGTACGATCCAGACAGTATCGCGGAGCGGACGAGCATGT 1300
QY      377 IleAspThrGlnGluAlaProLysThrProAlaProAlaAsnLeuGlnGluIleLys 396
DB      1301 AG-CCGCGCTCGGAATAATTCACGT---GAGCGCGTTCAACGCGATCGCTGTGATGCA 1356
QY      397 AlaGly 398
DB      1357 ATCGGA 1362

```

```

RESULT 27
BUJ2516      2147 bp  DNA      linear  BCF 04-DEC-2001
LOCUS      Bradyrhizobium japonicum nifK gene.
DEFINITION  AJ002516
ACCESSION  AJ002516
VERSION    AJ002516.1 GI:2597837
KEYWORDS   nifK gene; respiratory nitrite reductase.
SOURCE     Bradyrhizobium japonicum

```

```

ORGANISM  Bradyrhizobium japonicum
REFERENCE 1
AUTHORS   Velasco, L., Mesa, S., Delgado, M.J. and Bedmar, E.J.
TITLE     Characterization of the nifK gene encoding the respiratory,
          Cu-containing nitrite reductase of Bradyrhizobium japonicum
JOURNAL   Biochim. Biophys. Acta 1521 (1-3), 130-134 (2001)
MEDLINE   PUBMED 11690645
REFERENCE 2 (bases 1 to 2147)
AUTHORS   Bedmar, E.J.
TITLE     Direct Submision
          Submitted (30-OCT-1997) Bedmar E.J., Microbiologia del Suelo y
          Sistemas Simbioticos, Consejo Superior de Investigaciones
          Cientificas, CSIC, Estacion Experimental del Zaidin, P.O. Box 419,
          E-18080, Granada, Spain
FEATURES  Location/Qualifiers
          source          1..2147
                        /organism="Bradyrhizobium japonicum"
                        /mol_type="genomic DNA"
                        /strain="USD110"
                        /db_xref="taxon:375"
          gene            815..1992
                        /gene="nifK"
                        /bound_moiety="Fnr"
          protein_bind     884..889
                        /gene="nifK"
          RBS              898..1992
                        /gene="nifK"
          CDS              898..1992
                        /codon_start=1
                        /transl_table=1
                        /product="respiratory nitrite reductase"
                        /protein_id="CA05506.1"
                        /db_xref="GI:2597838"
                        /db_xref="SPTREMBL:O31380"
                        /translation="MPTRRALISAATATMLATPALADDLKLPKQVELVAPFVH
                        AHDQKQPKIVEFKLTIEKKVVIDEGTTFQMTFPGSMGPIVMVHESGVET
                        LNPATNTPPHNIDFSATGALGGALTLINPEQVYLKRAKTGTVFYHCGRGM
                        LPHVIVSGNAGAVMLPRDGLNDGKHALKYDVLVYVGQDMVPRDEKGNFYSYSP
                        GEASDTIEEMKKLIPSHVFNKGALGKALITNVGENVLIVASQANRDSRPHLI
                        GGHGVETGTEKFNAPVEGLTWFIRGSGAGAAVYKFMQPGIYAVVTNLILBADLC
                        ATAHFKVEKMDMLMTQVKAPREIPANTN"
          BASE COUNT      433 a 698 c 673 g 343 t
          ORIGIN
          Alignment Scores:
          Pred. No.:      3,346-20
          Score:          436.50
          Percent Similarity: 43.07%
          Best Local Similarity: 31.17%
          Query Match:    16.62%
          DB:              1
          Gaps:            16

```

```

US-10-088-045-2 (1-502) x BUJ2516 (1-2147)
QY      14 AlaLeuSerAlaLeuMetLeuSerGlyCysSerAsnGlnAlaAspLysAlaAlaGlnPro 33
DB      931 CCGGCGACCGCCCTGTATGCTGGCGACACCGCGCTGCGCGCGCATCTCAAGTCCG 990
QY      34 LysSerSerThrValAspAlaIleAlaLysThrAlaAsnAlaAspAlaIleSerGln 53
DB      991 CGCCAGAAAGTGAGCTGTGGCGCCCGCTTGTGATGCGACGACGAGGAGCAAG 1050
QY      54 GluHisGlnGlyGluLeuProValIleAspAlaIleValThrHisAlaProGluValPro 73
DB      1051 -----CAGGGC----- 1056
QY      74 ProProValAspArgAspHisProAlaLysValValLysMetGluThrValGlu--- 92
DB      1057 -----CCCAAGATGCTGAGTTCAAGCTCAAGCTCAAGATCGAGAG 1092

```

Qy 93 ---LysValMetArgLeuAlaAspGlyValGluTyrGlnPheThrPheGlyGlyGln 111  
 Db 1093 AAGAAAGGTCGTCATCGACGAGAGGATACCACTTCAGGGGAGACCTTCAAGGGCTCG 1152  
 Qy 112 ValProGlyGlnMetIleArgValArgGluGlyAspThrIleGluValGlnPheSerAsn 131  
 Db 1153 ATGCCGGGCGCGCTGATGCTGTCATGAGCAGCTACGCGAAGCGACCGCTGTCAT 1212  
 Qy 132 HisProAspSerLysMetProHisAsnValAspPheHisAlaIleThrGlyProGlyGly 151  
 Db 1213 CCGCGACCAACACCATGCGCGACAACTGCTTCACTTCGCGACCGGGCCCTGGC 1272  
 Qy 152 GlyAlaGluAlaSerPheThrAlaProGlyHisThrSerThrPheSerPheValAlaLeu 171  
 Db 1273 GGGCGCGCGCTCACCTGATCATCCGCGAGAGGTCGTCGCGCTGGAGAGCGGACC 1332  
 Qy 172 GlnProGlyLeuTyrValTyrHisCysAlaValAlaProValGly-----Met 187  
 Db 1333 AAGACCGCGCGCTGCTCTATCACTGC-----GCGCGCGCGCGCGCGATGATCCCTGG 1386  
 Qy 188 HisIleAlaAsnGlyMetTyrGlyLeuIleLeuValGluProLysGluGlyLeuPro--- 206  
 Db 1387 CAGCTGCTCTCCGATGACGAGCGTGCCTGATGCTGCGCGCGAGCGAGCTAAGCAGAC 1446  
 Qy 207 -----LysValAspLysGlyTyrTyrValMetGlnGlyAspPheTyr 220  
 Db 1447 GGCAAGGCGCACGCGCTGAATATAGACAGCTCATATGCGCGAGAGAGCATATAC 1506  
 Qy 221 Thr-----LysGly-----LysTyrGlyGluGlnGlyLeuGlnProPhe 233  
 Db 1507 GTGCCGCGCGACGAGAGGCAATTTCAAGTCTATACCTCGCGCGGCGAAGCTTCACC 1566  
 Qy 234 AspMetGluLysValAlaIleArgGlu---AspAlaGluTyrValAlaPheAsnGlySerVal 252  
 Db 1567 GACACCGAAGAGATGAGAGAACCTGATCCCATGCTGCTGCTGCTCAAGGAGAGCTC 1626  
 Qy 253 GlyAlaLeuThrGlyGluAsnAlaLeuLysAlaLysValGlyGluThrValArgLeuPhe 272  
 Db 1627 GGGCGCGCTTACCGCGCAAGATGCGCTCACCGCAATGCGCGAGAAAGCTG-----CTG 1680  
 Qy 273 ValGlyAsnGlyGlyProAsnLeuThrSerSerPheHisValIleGlyGluIlePheAsp 292  
 Db 1681 ATCTGTCACCTCGCAGCGCAATCGCGACAGCGCGCGCATGATGCGCGCGCATGCGGAC 1740  
 Qy 293 LysValHisPheGlyGlyGlyGlyGluAsnHis-----AsnIleGlnThrThr 309  
 Db 1741 TATGTC---TGGGAGACCGGCAAAATTGGGCAATGACCCGAGGTCGGGCTGGAGACCTGG 1797  
 Qy 310 LeuIleProAlaGlyAlaAlaIleThrGluPheLysValAspValProGlyAspTyr 329  
 Db 1798 TTATCCGGCGCGCGCTGCGAGCGGCGTCGATGACAAATTCAGCAGCCCGGATATAC 1857  
 Qy 330 ValLeuValAspHisAlaIlePheArgAlaPheAsnLysGlyAlaGluGlyIleLeuLys 349  
 Db 1858 GCGTATGTCACGCTAATCGATGAGAGCGCGCGCATCTGCGCGACCGCATCTTCAAG 1917  
 Qy 350 ValGluGlyGluGluGluAsnHisGluIleTyrSerHisLysGlnThrAspAlaValTyrLeu 369  
 Db 1918 GTGGAAGGCAAGTGGAAGCAGCATCTGATGACCAAGGTGAAGCGGCTGCG----- 1968  
 Qy 370 ProGluGlyAlaProGlnAlaIleAspThrGlnGluAlaProLysThrProAlaProAla 389  
 Db 1969 -----GAAATCCCGCGCAACACC 1986  
 Qy 390 AsnLeuGlnGluGlnIleLysAlaGlyLysAlaThrTyrAspSerAsnCysAlaAlaCys 409  
 Db 1987 AACTAGACCGCAGCAA---GGGCGCGCGG----- 2010  
 Qy 410 HisGlnProAspGlyLysGlyValProAsnAlaPheProProLeuAlaAsnSerAspTyr 429  
 Db 2011 -----GGAATCACCGGCGCCCTTC---TTCTTTGCGGGGAGAGCATG 2049

Qy 430 LeuAsnAlaAspHisAlaArgAlaAlaSerIleValAlaAsnGlyLeuSerGlyLysIle 449  
 Db 2050 CTGATGCACTTCAAGCTGAAGATGTCGCTGCTGCGCGCGAGCTCGCAGAGCGCATC 2109  
 Qy 450 ThrVal 451  
 Db 2110 GCGGTG 2115  
 RESULT 28  
 AE009732/c 11630 bp DNA linear BCT 20-MAR-2003  
 LOCUS Brucella melitensis 16M chromosome II, section 94 of 107 of the  
 DEFINITION complete sequence.  
 ACCESSION AE009732 AE008918  
 VERSION AE009732.1 GI:17985201  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Brucella melitensis 16M  
 Brucella melitensis 16M  
 Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 Brucellaceae; Brucella.  
 1 (bases 1 to 11630)  
 DelVecchio, V.G., Kapral, V., Redkar, R.J., Patra, G., Mujer, C.,  
 Los, T., Ivanova, N., Anderson, I., Bhattacharya, A., Lykidis, A.,  
 Reznik, G., Jablonski, L., Larsen, N., D'Souza, M., Bernal, A.,  
 Mazur, M., Goldsman, E., Selkov, E., Elzer, P.H., Hagius, S.,  
 O'Callaghan, D., Letesson, J.-J., Haselkorn, R., Kyrides, N. and  
 Overbeek, R.  
 The genome sequence of the facultative intracellular pathogen  
 Brucella melitensis  
 Proc. Natl. Acad. Sci. U.S.A. 99 (1), 443-448 (2002)  
 11756688  
 2 (bases 1 to 11630)  
 DelVecchio, V.G., Redkar, R.J., Patra, G. and Mujer, C.  
 Direct Submission  
 Submitted (13-NOV-2001) Institute of Molecular Biology and  
 Medicine, University of Scranton, Scranton, PA 18510, USA  
 3 (bases 1 to 11630)  
 Elzer, P.H. and Hagius, S.  
 Direct Submission  
 Submitted (13-NOV-2001) Department of Veterinary Science, LSU Ag  
 Center, 111 Dairymple Building, Baton Rouge, LA 70803, USA  
 4 (bases 1 to 11630)  
 Kapral, V., Los, T., Ivanova, N., Anderson, I., Bhattacharya, A.,  
 Lykidis, A., Reznik, G., Jablonski, L., Larsen, N., D'Souza, M.,  
 Bernal, A., Mazur, M., Goldsman, E., Selkov, E., Haselkorn, R.,  
 Kyrides, N. and Overbeek, R.  
 Direct Submission  
 Submitted (13-NOV-2001) Integrated Genomics, Inc., 2201 W. Campbell  
 Park Drive, IL 60612, USA  
 5 (bases 1 to 11630)  
 Letesson, J.-J.  
 Direct Submission  
 Submitted (13-NOV-2001) Unite de Recherche en Biologie Moleculaire,  
 Laboratoire d'Immunologie et de Microbiologie, Universite of Namur,  
 61 rue de Bruxelles, Namur 5000, Belgium  
 6 (bases 1 to 11630)  
 O'Callaghan, D.  
 Direct Submission  
 Submitted (13-NOV-2001) Faculte de Medecine, INSERM U431, Avenue  
 Kennedy, Nimes 30900, France  
 Location/Qualifiers  
 1. 11630  
 /organism="Brucella melitensis 16M"  
 /mol\_type="genomic DNA"  
 /strain="16M"  
 /db\_xref="taxon:224914"  
 /chromosome="II"  
 /complement(132..1262)  
 /gene="BMEI10988"  
 /complement(132..1262)  
 /gene="BMEI10988"  
 /FC\_number="1.7.99.3"  
 /codon\_start=1

```

/translate=1
/product="COPPER-CONTAINING NITRITE REDUCTASE PRECURSOR"
/protein_id="AAL54230.1"
/db_xref="GI:17985202"
/translation="MADQIQVNRRTILAGALGCVLSATSAMGCGTMRKAAEE
IALLPRKVELVPPFVHATQVABGSPKVVQTPWTEKKYIIDAGTEVHAMTRNG
TVGSLPMVHQDDYLETILNPSTNTLHMDIDPAHALGAGGGLTEINKEGTIVLRF
KATKPGVYVHCAPEGVPMWVSGMGAIVMLPREGLHDGKNKLTLYDKVYVGEDD
FYVPRDNGNYKTYEAGDAEDTVKVMRTLTPTHVFNAGVALTGDKALTFKVGESK
VLIIISOANDRTREPLHIGHDYVMATGKFTPPDVQETWTFIPGAAAAAFTFRQP
GIYAVYVNHNLIEAFELGAAPFKVTGEMNDLMTSILAPSGM"
complement (1618. .2220)
/gene="BME110989"
/complement (1618. .2220)
/gene="BME110989"
/codon_start=1
/translate=1
/product="hypothetical protein"
/protein_id="AAL54231.1"
/db_xref="GI:17985203"
/translation="MGLLAGALVLPSPAPAKTQQAAMPNATSPHQADVLLRGA
DIRSTGIDEGELQAGVAVNAHVGAAHMLVNRIVADQCKGLPVLIGSLGN
AATVIAEELERGLIADVMAFPAATGPDPLPGVVRVYVNFYFKQHGKGLPVGPRFH
GHLENDPSNAKVGHPNTEKQRPLOAEVVDVLAVVNAD"
complement (2267. .2851)
/gene="BME110990"
/complement (2267. .2851)
/gene="BME110990"
/codon_start=1
/translate=1
/product="farnesyl cysteine carboxyl-methyltransferase"
/protein_id="AAL54232.1"
/db_xref="GI:17985204"
/translation="MMPLALIVMSIGLVAVVIRIPHORARKIKVARTATLADRL
ALCAATVGLSVPLVYVATGFPKADTFPPMGMIGTILELFLFVASHQNGN
WSYSLDIRREHKLVTDLRYVRHPMLTSLWLAIQFLLPMMIAGLGLVALITY
FYRIEHEAEMRAFGSAYDEYSSRIGRIIPKPM"
complement (2848. .4023)
/gene="BME110991"
/complement (2848. .4023)
/gene="BME110991"
/codon_start=1
/translate=1
/product="Hypothetical Membrane Spanning Protein"
/protein_id="AAL54233.1"
/db_xref="GI:17985205"
/translation="MQHPVRRRAIFPGYDPTPEAFENRLREIRFAGALDVKTQ
VSPVSEVASSSEIGSVITETRSAGWSYQTFNPLVDKIVLAFARPLPRLMKYVLA
FADFVLSGTAFFEFARARFALFYLPFLALVFLAGMLAARLTAPWLIGAMIVGL
FVFFPLATLTGNMSTNHLMDLMSFSNFRGORDPAELMORFAEIDRVAAGHD
EVIVIGSTGCMMLDVARCLADRAFPATRPVVLITLGSALAKAGYHPAAGKPO
DNORLVDPEALFENCEVOCLTDVNFYTFDVAEMKLRPADRAFPVROVRITDMOP
DIYRIKRNFRVHYQVFGNTKAYWYDFQICGPAISLSERMMDLVGTAKESAQQ"
complement (4147. .4473)
/gene="BME110992"
/complement (4147. .4473)
/gene="BME110992"
/codon_start=1
/translate=1
/product="WRP PROTEIN HOMOLOG A"
/protein_id="AAL54234.1"
/db_xref="GI:17985206"
/translation="MEBKOLERIKACVREALRAVODEPLAVNLVDLGMIVAVEISDQ
MNVKMTTTRGCPAAGFLQAVQACIEIEIGTGAARVELTYEPENKPEMAITEVQAI
FAAPSP"
complement (4480. .5007)
/gene="BME110993"
/complement (4480. .5007)
/gene="BME110993"
/codon_start=1
/translate=1
/product="hypothetical protein"
/protein_id="AAL54235.1"

```

```

/db_xref="GI:17985207"
/translation="MTVOELAPYEILDVRPLSGVYRFOALINGAVDSLAPGQGLLIIA
PEKPYLLFVAILERKGEAPKEPIEGEDMOVLFSPPVPEPLAEVQLSKNVSGPDITWPL
OYLDCTDMDPPEPLVYLIAEVENMSAGEVILFALIHREPLFLPELENRGHQWGNFDE
TKQYRIMLRVGGCG"
complement (5004. .5840)
/gene="BME110994"
/complement (5004. .5840)
/gene="BME110994"
/codon_start=1
/translate=1
/product="hypothetical protein"
/protein_id="AAL54236.1"
/db_xref="GI:17985208"
/translation="MLNITAGLSAPYIHAALGLGMMFTFAMGVSYRLTMTFLSPDE
RRTYRLVMMWAGAALTVLAGVAFAPFGODMADVLLAALLPGVASVLYAFPMRAVY
ROKRRKATILNSAASIPAPAAVIALILALAPMCPSPMIGALVYLPFGMLTGS
LAOLYIVFPLTLECYGVMGRVPTPRQDIYSEKPARMFCITYTGVALATLALGA
GYPVAFQLASALNLIALITVHFFRARRLMDVEAVRLPENGVAIPHLIYAGGVPPRR
SK"
complement (6407. .6721)
/gene="BME110995"
/complement (6407. .6721)
/gene="BME110995"
/codon_start=1
/translate=1
/product="hypothetical protein"
/protein_id="AAL54237.1"
/db_xref="GI:17985209"
/translation="MDGPRARCQVNLFPPLUSSTPAGPPPTTHATITSNIDTLE
PGSEPIVNDHDPVPLRNQLEHRNPGAFSVMYKETPLRVEIGRRKVAHSDHECTG
NH"
complement (6878. .8779)
/gene="BME110996"
/complement (6878. .8779)
/gene="BME110996"
/codon_start=1
/translate=1
/product="NORD protein"
/protein_id="AAL54238.1"
/db_xref="GI:17985210"
/translation="MLDPLELEETVGRANRHLIGKTSWPOYDHAVALVDIQRLLAI
CFRFGGDIAYVOIAPARARTSHRLGRMALGEBELQPLDEATMLMPELALFP
DROLNDLYLVMLGVYAVAMPDADALPELARLDLALQIAEQTVRACAPFGLKPR
YKRLCAAILAERPKRPLHRLQEOVEARLISIKOGADLPDADLPTTFPHGSPAGYLA
LPVLPMPGKMKREYVAPRGCDEPVPVNSQSGEGRQIAORRORPRHADRPETLN
RREKTIAMAMVSVDRPSGSDONAKSADIEDDLTUGERKGRPAARFRFDLDPBEA
VBSLLTALBLYPEMDYRGAVALPNHCAVLAAGVQEKPELEDAAGSIVRVRRQF
EILRRPEREYLRQLDGTDLDLDAVVRSCDLAAGGSGSDRVHLSRPOANDLAVTLLV
DVSLESTDADVNRVLDVEKEALLVLANGIAACGDRCSILTFSTRRSWVRVETVDF
DESFGPTVHRIALKRGFTYTRGAMRHATLAKAOPNRKKLLILLTGCKPRDVHY
EGRFALDEDRRAAGVFRAGVAVFATTVDBEASAVYLPALFGRGYALVANLALPVAL
PATRYVLAC"
complement (8791. .9603)
/gene="BME110997"
/complement (8791. .9603)
/gene="BME110997"
/codon_start=1
/translate=1
/product="NOKO PROTEIN"
/protein_id="AAL54239.1"
/db_xref="GI:17985211"
/translation="MSSLKAVETASPAIPAYSPSGNECALFEMAMORKLPLLLKQPT
GCKKTRFVHMAARMOLPLSTVTSCHDDLAADLTGRYLLKGGOTTIMDDGLTFAVVDG

```

```

Alignment Scores:
Pred. No.: 5.33e-19
Score: 431.00
Percent Similarity: 48.43%
Best Local Similarity: 34.76%
Query Match: 16.41%
DB: 1
Length: 11630
Matches: 122
Conservative: 48
Mismatches: 147
Indels: 34
Gaps: 11

```







Db 5718 GATCTCCCTTCGTCTCATGCA-----CACATCAAGTTCGCCGAAGGGCGCCGAAG 5768  
Oy 83 LysValIValIValIYmeGluThrValIGluYValIMetArgLeuAlaAsp--GlyVal 101  
Db 5769 GTGGTTCAAGTTACCATGGATGATCGAGGAAAAAGAAAGATCGTCATGATGAATGCGGGCAGC 5828  
Oy 102 GLuTYrGLInPhETrPTHrPheGlyGLyGLInValIProGlyGLInMetIleArgValArgGlu 121  
Db 5829 GAAGTCCATGCCATGACTTTCACCGAGGACGGTGGCCCGGACCGCTCGTGTGTGTGCATCAG 5888  
Oy 122 GLysPThrIleGluValIGlnPheSerAsnHisProAspSerLYmeProHisAsnVal 141  
Db 5889 GACGATTATCTGAACTGACAGCTCATTTAAACCCGGAACCAATACCTCTTGACAAACATC 5948  
Oy 142 AsPheHisValIAlaIleThrGlyProGlyGLyGLyAlaGluAspPheThrAlaProGly 161  
Db 5949 GATTTCATGCGCGCAACCGCGCACTGGGTGGTGGCGGGCTGACGAGATCAATCCGGGC 6008  
Oy 162 HisThrSerThrPheSerPheLYmeValAlaLeuGlnProGlyLYmeTYrValTYrHisGlyAla 181  
Db 6009 GAGAAAGACCTTCGCTCGCTTCAAAACCAACCAACTGGCGCTTTGTGTATCACAGC--- 6065  
Oy 182 ValAlaProValIGlyMet-----HisIleAlaAsnGlyMetTYrGlyLeuIleLeu 198  
Db 6066 ---GACATCCCGGAATGGTTCCTCGGCACGTGTATGTGGCATGAACGGCGCTGTCTGT 6122  
Oy 199 ValGluProLYmeGluGlyLeuPro-----LYsValAspLYeGlu 211  
Db 6123 GTTTCGCGCGCGAGGGGTGGTCATGACGGCAAGGGCAACAGTCACTATGACAAAGTCT 6182  
Oy 212 TYrTYrValIMetGlnGlyAspPheTYrThrLYs-----GlyLYs 224  
Db 6183 TATTATGTCTCGGACAGAGATTTCTATGTCCGCGCTGATGAGAACGCAATTACAGAC 6242  
Oy 225 TYrGLyGluGlnGlyLeuGlnProPheAspMetGluYsValIleArgGlu--AspAla 243  
Db 6243 TATGAAGCGCGCGGAGTCCCTATGAGATACGCTCAAGGTGATGCGACCTGACGCCG 6302  
Oy 244 GLuTYrValIValIAsnGlySerValGlyAlaLeuThrGlyGLyGLInAsnAlaLeuLYsAla 263  
Db 6303 ACCCATGTGTCTTCAATGATGTGGGTGGTGTCTTACGGGTGACAAAGCCCTGACGGCG 6362  
Oy 264 LYsValIGlyGluThrValArgLeuPheValIGlyAsnGlyGLyProAsnLeuThrSerSer 283  
Db 6363 AAATCGCGGAAAGAGTGTCTGATCTC-----CACTGCAAGGCCAACCGGATACCCGT 6416  
Oy 284 PheHisValIleGlyGluIlePheAspLYsValHisPheGluGlyLYsGlyGluAsn 303  
Db 6417 CCGCACTCGATCGCGCGCGCATGCGCATATGTCTGGGCGACCGGCACTCATACGCCG 6476  
Oy 304 HisAsnIle-----GlnThrThrLeuIleProAlaGlyGLyAlaAlaIleThrGluPhe 321  
Db 6477 CCAAGTCTGACACAGAGAACTCGTCTCATCTCCGGCGGTGGCGACCGCGCACTTCAC 6536  
Oy 322 LYsValAspValIProGlyAspTYrValLeuValAspHisValIlePheArgAlaPheAsn 341  
Db 6537 ACGTTCCGGCAACCGGGCATCTACCGCTATGTCATCTAATCACTGATGAAAGCTTTGAG 6596  
Oy 342 LYsGlyAlaIleGluGlyIleLeuLYsValIGluGlyGluGluAsnHisGluIleTYrSerHis 361  
Db 6597 CTGGGCGCGGTGGCCCACTTCAAGGTGACGGGCGAATGAGATGACAGACTTGATGACA--- 6653  
Oy 362 LYsGlnThrAspAlaValTYrLeuProGluGly 372  
Db 6654 -----TCAATACCTGGCGCGCTCGGGA 6674

RESULT 30  
BX321859/c BX321859 316050 bp DNA linear BCT 23 -APR-2003  
LOCUS Nitrosonomas europaea ATCC 19718, complete genome; segment 4/10.  
DEFINITION BX321859 AL954747  
ACCESSION  
VERSION BX321859.1 GI:30138466

KEYWORD	complete genome.
SOURCE	Nitrosomonas europaea ATCC 19718
ORGANISM	Nitrosomonas europaea ATCC 19718
AUTHORS	Bacteria; Proteobacteria; Betaproteobacteria; Nitrosomonadales; Nitrosomonadaceae; Nitrosomonas.
REFERENCE	1 Chain, P., Lamerdin, J., Larimer, F., Regala, M., Land, M., Hauser, L., Hooper, A., Klotz, M., Norron, J., Sayavedra-Soto, L., Arciero, D., Hommes, N., Whitaker, M., and Arp, D. Complete Genome Sequence of the Ammonia-Oxidizing Bacterium and Obligate Chemolithoautotroph Nitrosomonas europaea J. Bacteriol. 185 (9), 2759-2773 (2003)
JOURNAL	12700255
PUBMED	2 (bases 1 to 316050)
REFERENCE	Larimer, F.
AUTHORS	Direct Submission
TITLE	Submitted (12-NOV-2002) Submitted on behalf of the Nitrosomonas genome consortium, the DOE Joint Genome Institute, Production Genomics Facility, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA, and the Genome Analysis Group, Oak Ridge National Laboratory, 1060 Commerce Park Drive, Oak Ridge, TN 37831, USA;
JOURNAL	larimerf@ornl.gov
REMARK	Nitrosomonas genome consortium
FEATURES	Location/Qualifiers
SOURCE	1. 316050
	/organism="Nitrosomonas europaea ATCC 19718"
	/mol_type="genomic DNA"
	/strain="ATCC 19718"
	/db_xref="taxon:228410"
	/complement (639..1223)
	/locus_tag="NE0846"
	/complement (639..1223)
	/locus_tag="NE0846"
	/codon_start=1
	/evidence=not experimental
	/transl_table=11
	/product="conserved hypothetical protein"
	/protein_id="CAD84757.1"
	/db_xref="GI:30138467"
	/translation="MIRNNSEFYRIDPLKPIKANPLRNGESLNHRLQNIPLDGLIVPYFLALCEFLIAMEWIRMYTOTSPNPVLFYTMALPCITLMLKTYKGEYKRIKLGLGEALVGOFLERLRACGSHIFHDIPEGKFNLDVHVHSSGIYVETKTLSPDRGESKLVYGNHNLTKNSLALDRNPITOVNRNSWLE"
	order(complement(1038..1094),complement(951..1007))
	/locus_tag="NE0846"
	/note="2 probable transmembrane helices predicted by TMHMM2.0"
	/complement (1351..2301)
	/locus_tag="NE0847"
	/complement (1351..2301)
	/locus_tag="NE0847"
	/function="COGS COG0726"
	/codon_start=1
	/evidence=not experimental
	/transl_table=11
	/product="conserved hypothetical protein"
	/protein_id="CAD84758.1"
	/db_xref="GI:30138468"
	/translation="MKLIPSLFSPAGKNSKLISIVYHRLPEPDLTLGEGIAKOFEGKGLSVLTNNFLLPLSEANVTKLRSGTLPGRACITPDGCVADNARIALPILOKHGTAFPLFSGPFDGGRMNDYIVESVRPAKGDKDLNLAIGLGNALATLEQRRETLNLLINKLKYLPHEHRAQVDKSLTIIADLPDLNLMISAQVROLDHGMETIGGHTLTHPLIADSDAARTEIAIGAKALEAIIIDTPIRLFLFYPPGKGQKYLASHVMKIKDAGEAAVSTAMGAARODSMFQLPRFTPMWDSSEWRVYVLRMVRNRRRIQTV"
	complement (2421..2873)
	/locus_tag="NE0848"
	/complement (2421..2873)
	/locus_tag="NE0848"
	/function="Interpro IPR001345 COGS COG2062"
	/codon_start=1
	/evidence=not experimental
	/transl_table=11
	/product="Phosphoglycerate mutase family"

```

gene
CDS
    /protein_id="CAD84759.1"
    /db_xref="GI:30138469"
    /translation="MDLILMRHAEADGIPDARTRKLTENGLKQAKKARWLEPKLPD
    TRIIVPATRQOTVSGALTHFETSEQVGSATPHRVANTIAMPEAGCTVVLVGHQPT
    LKRIASLLKGDGSGFSVRKSGSIWMFSSKQKDERESILLRAVMTPELL"
    complement(2929..3621)
    /locus_tag="NE0849"
    /locus_tag="NE0849"
    /locus_tag="NE0849"
    /function="COG30814"
    /codon_start=1
    /evidence=not_experimental
    /transl_table=11
    /product="similar to nodulin 21"
    /protein_id="CAD84760.1"
    /db_xref="GI:30138470"
    /translation="MTHDNTYHSGRTGMLRAAVIGANDGIIVTSALILIGVASAHAAD
    DLLACVAGVYAAKMSKAGEYVSSQSTEDKADVALBEYHLDRIIDRIQLDTLIT
    MKGLOPELAAQVARELMADALDHLRDLGHERVNAKPVQAFTSAGMFTLGASM
    PLAAITAAPATTHIIPVAISSLSTALGTFAAYIGKAMLTGAARVAFWALAMAF
    TALGTLLFGIIA"
    order(complement(3496..3561),complement(3418..3483),
    complement(3112..3177),complement(3034..3099),
    complement(2935..2997))
    /locus_tag="NE0849"
    /locus_tag="NE0849"
    /note="5 probable transmembrane helices predicted by
    TMHMM2.0"
    complement(3637..4311)
    /locus_tag="NE0850"
    complement(3637..4311)
    /locus_tag="NE0850"
    /function="InterPro IPR000379:IPR003140 COGS COG0400"
    /evidence=not_experimental
    /codon_start=1
    /transl_table=11
    /product="Phospholipase/Carboxylesterase"
    /protein_id="CAD84761.1"
    /db_xref="GI:30138471"
    /translation="MPDNSFQLSAIDITTSNPBYTILMMHGLGADNDPEVVOALD
    LPEIPIRFLPEPAPQOVPTINGVIMRAWYDIOTDFEVEDGEGIRRSQHAIVELIE
    REDRGRIPDHLILAGSGOAMALHTGRLPHPRILAGITLISGLPLAKIREAHIT
    NRTPIFMAHGNDPIVPIELALASLQQLREYYPVTWHYPMEHVTCDELDVDSHW
    LKTLIK"
    4378..4458
    /note="Signal predicted by SignalP 2.0 HMM (Signal peptide
    probability 1.000) with cleavage site probability 0.844 at
    residue 27"
    4390..5502
    /locus_tag="NE0851"
    4390..5502
    /locus_tag="NE0851"
    /function="InterPro IPR002931 COGS COG1305"
    /codon_start=1
    /evidence=not_experimental
    /transl_table=11
    /product="Transglutaminase-like superfamily"
    /protein_id="CAD84762.1"
    /db_xref="GI:30138472"
    /translation="MKRRSFFKLVGSSAAALPFAVARSLEPPNOKMSYQUTYQVT
    LPSAGVAKMLPLPASDVTYQTOGNNNSGATYSAFTMLGCTOSQVFEALWQGE
    ERTVRVSSIVKTRDVTDLHSYRAPAKISIPONIORLOSTKRIPLDDVVKCARSVT
    SPNAHTPLQOARAIYDWINVAVDINGGCGDLRLSLSKRENISGCKGRIHALF
    VGLARASGITPARIOYGRIGDSALKNKLKIGDSISTSOHCOAFYISGLGWPVHNS
    VARVMAESLSHHSDISOLREKLFSGSWENWYAFNRRENVLDGTTKVSALISKLPEF
    GIPVAIDGRLODSLPEBSFSYKITSALVLTGAKL"
    complement(5615..7336)
    /gene="yv90"
    /locus_tag="NE0852"
    complement(5615..7336)
    /gene="yv90"
    /locus_tag="NE0852"
    /EC_number="1.8.1.2"
    /function="InterPro IPR000660:IPR005117 COGS COG0155"

```

```

gene
CDS
    /codon_start=1
    /evidence=not_experimental
    /transl_table=11
    /product="Nitrite and sulfite reductase"
    /protein_id="CAD84763.1"
    /db_xref="GI:30138473"
    /translation="MANTLPEDRSCDIOPLERLSPBDSLKAESDYLRGTLALGLLD
    RITSAPGNDIKLKMFKPIYEDDREIRBERRQKLEPAQFMIRVRLPGICTTERM
    LKISELACAGNETLMTTRQTFQFMVKONIVPLRGLHEVLDTVAACGDSRGV
    MATVNPOPALOELEALAKTVSDHVPKTRVAHEIWMYGERIASSPEEPGGOTYM
    PREKIGFVTPPNNDIDYADOLGYAIIIGENCKIGFNVATIGCGMRDABHTYPR
    TASVIGFTIPDRILISTEAVMGVQYDRGNRADSRRKFTITDDKGLDIKALIEDRA
    GPLESARPYDFPSNADIGYIESGDFHFTLLEENGRLNRLMDLIDIAQIAHVKHF
    RLTPNOLMIANVATADKPEIEALRETEGLIARNESVRLNSMACVALPTCLAMR
    SERVLPDLITKIEGLTRYNLQNEPITLMGTCGNCSPFIIEIGLGRAPKYNLY
    LGGCFHGRQNLNRYRENIEBPALLETLENLVGLRGAIVARLIPDEHFGFTIRAGIIRVET
    EGRFSD"
    complement(7345..9180)
    /gene="yv9r"
    /locus_tag="NE0853"
    complement(7345..9180)
    /gene="yv9r"
    /locus_tag="NE0853"
    /EC_number="1.8.1.2"
    /function="InterPro
    IPR001094:IPR001226:IPR001433:IPR001709:IPR003097 COGS
    COG0369"
    /codon_start=1
    /evidence=not_experimental
    /transl_table=11
    /product="Sulfite reductase flavoprotein subunit"
    /protein_id="CAD84764.1"
    /db_xref="GI:30138474"
    /translation="WMAQSLSRGLTJLABOMOLVEALIGSSLTPEHAWYISGFAGLGA
    LRRSGTASQSPDVCIPIPVYTOABEPAAASLTILHISRTGNSFELAIRLAALVEGL
    SPTLVGIADYKVRKLEBODLLITSTHSGBDPQSGMEFEEVERKASLALRAA
    IIALGMSYEHFQAGKRLDEREALGAARLQRPVDCVDEDPAAVWSGILALLAA
    BQSAISVYASPSKTTGTQOANSYVSRNPPAVINDIVLSGSGSTETHHIIS
    LBDGSLTVEGALGIAANDPAMVALLAALMNDAPLTVGQOTTLPLDAGKSE
    ITTATLRFPHMGLTGTIDPSAMNTRAAFLHNHIIIDIRAYPLKGLDORFVD
    GLRPDOPRIYSLASSISTPDEAHITTAIVRYLKHSPRSVAGSHLADPAIDSVLP
    VTIQSPHRLPSNGEPIITMGATVAVPRAVWOREANAGGRSMTFERNPRIDF
    LYQTEPDWMLKNGVLTAKMEVAPSRODAEVYVQHRITYEAHADVYAMLLEGAVIYLCGD
    ASHMAPDHNILVLTVEEQASVQREAEERYLRDLQRDGRYQRDVY"
    complement(9708..10664)
    /gene="cysB"
    /locus_tag="NE0854"
    complement(9708..10664)
    /gene="cysB"
    /locus_tag="NE0854"

```

Alignment Scores:

Pred. No.:	3,366-17	Length:	316050
Score:	428.50	Matches:	104
Percent Similarity:	54.81%	Conservative:	44
Best Local Similarity:	38.52%	Mismatches:	98
Query Match:	16,32%	Indels:	24
DB:	1	Gaps:	6

```

US-10-088-045-2 (1-502) x BX321859 (1-316050)
Qy 100 GYVAIGLTYGCGINPHERPPTRRPHEGLYGLVGLNVALPQGLYGLINMETLEARGVAL 119
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 88010 GGAAGTACTTATGTCGCTCGAGCTTCGACGCGTAAGCTTCCCGTCGACGCGTC 87951
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Qy 120 ARGGLUGLYASPTHTLLEGLVALGINPSESASNHIEPRQASPERLYSWETPROHIE 139
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 87950 ACTGAAGGGATACGGTGAATTCATCCATCAATGACAAATAATGTAAGAATCCCAT 87891
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Qy 140 ASNVVALASPHETISLALALAL--THRGCLYPROGLYGLYGLYALAGLUALASERPHETNR 158
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 87890 TCCATGATTTCCATGCTCGACGCGTGAATGCTGAGAGATTTCCAGTCTG-----ATC 87837
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||

```

Qy 159 AlaProgluYhiSerThrPheSerPheValAlaLeuGlnProgluYLeuTyrValTyr 178  
 Db 87836 AAGCGGCTGAACAAAAAATACCTTACTGATATCCCGGCTGATTTTAT 87777  
 Qy 179 HisCysAlaValAlaProValGlyMetHisIleAlaGlnGlyMetTyrGlyLeuIleLeu 198  
 Db 87776 CACGTGGCTGTGATCCGATATTCACACATCCACGATATGTCGATGATATC 87717  
 Qy 199 ValGluProGlyGlu-----GlyLeuProGlyValAlaGlyTyrValMetGln 216  
 Db 87716 GTGATCCGAAAGATGCAATGATACCCAGGACGATCTGATATGCTGATTCAG 87657  
 Qy 217 GlyApePheTyrThrLysGlyLysTyrGlyGlnGlnGlyLeuGlnProPheAspMetGlu 236  
 Db 87656 GCTGAGCTTACGAG-----AATCTGACGACAAAAA 87624  
 Qy 237 LysAlaIleArgGluAspAlaGluTyrValValPheAsnGlySerValGlyAlaLeu--- 255  
 Db 87623 GCCATGATGCAACAAATGAGATATGTCCTATTCATGATGAGGGGTTTCAAGTATGAC 87564  
 Qy 256 -----ThrGlyLysAlaLeuLysAlaLysValGlyLysValArg 270  
 Db 87563 CCGGTGATGATTTCTGAAGCCACCAAGCTGCTGACGCCAAACAGGTGAACGGGTGCGC 87504  
 Qy 271 LeuPheValGlyAsnGlyLysProAsnLeuThrSerSerPheHisValIleGlyLysIle 290  
 Db 87503 ATTACTTCGTCACACGCCGACCAATGATTTGATCTCTTCAACCCCATGACAGGATC 87444  
 Qy 291 PheAspLysValIlePheGlnGlyLys-----LysGlyGluAsnHisAsnIleGlnThr 309  
 Db 87443 TGGATCGGGTTTACCCACGCGGCAACCCCAAAACGTCAGTATGCTGCAATCTCTAT 87384  
 Qy 310 LeuIleProAlaGlyLysAlaAlaIleThrGluPheLysValAspValProGlyAspTyr 329  
 Db 87383 CTGATAGAGCGGAGATGCGGCAACGCTGATCTGATTTCTCCGTTGAAGGTGCCAAC 87324  
 Qy 330 ValLeuValAlaPheHisAlaIlePheArgAlaPheAsnLysGlyAlaLeuGlyIleLeuLys 349  
 Db 87323 GCGATCGTCATCTGATGATG---CGTCATGACACATCCGTCGATGCGCATATATG 87267  
 Qy 350 ValGluGlyGluGluAsnHisGly 357  
 Db 87266 TTCATTAATGACGCTGATCCGAA 87243

RESULT 31  
 LOCUS AFANIR 2351 bp DNA linear BCT 03-FEB-1999  
 DEFINITION A faecal gene for nitrite reductase, complete cds.  
 ACCESSION D13155  
 VERSION D13155.1 GI:398121  
 KEYWORDS nitrite reductase.  
 SOURCE Alcaligenes faecalis  
 ORGANISM Alcaligenes faecalis  
 Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
 Alcaligenaceae; Alcaligenes.  
 REFERENCE 1 (bases 1 to 2351)  
 AUTHORS Nishiyama, M., Suzuki, J., Kukimoto, M., Ohnuki, T., Horinouchi, S. and  
 Bapnu, T.  
 TITLE Cloning and characterization of a nitrite reductase gene from  
 Alcaligenes faecalis and its expression in Escherichia coli  
 JOURNAL J. Gen. Microbiol. 139 (Pt 4), 725-733 (1993)  
 MEDLINE 93294530  
 PUBMED 8515232  
 2 (bases 1 to 2351)  
 AUTHORS Nishiyama, M.  
 TITLE Direct Submission  
 JOURNAL Submitted (01-SEP-1992) Makoto Nishiyama, The University of Tokyo,  
 Department of Biotechnology, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113,  
 Japan (Tel:03-3812-2111 (ex.5126), Fax:03-3812-0544)  
 COMMENT Submitted (01-SEP-1992) to DDBJ by:  
 Makoto Nishiyama  
 Dept. of Agricultural Chemistry  
 Faculty of Agriculture

FEATURES  
 source  
 The University of Tokyo  
 1-1-1 Yayoi, Bunkyo-ku  
 Tokyo 113  
 Japan  
 Phone: 03-3812-2111 x 5126  
 Fax: 03-3812-0544  
 Location/Qualifiers  
 1. .2351  
 /organism="Alcaligenes faecalis"  
 /mol\_type="genomic DNA"  
 /strain="S-6"  
 /db\_xref="taxon:511"  
 misc\_binding  
 /bound\_moiety="putative FNR binding site"  
 RBS  
 737..743  
 751..1881  
 /gene="nir"  
 751..1881  
 /gene="nir"  
 /codon\_start=1  
 /transl\_table=11  
 /product="nitrite reductase"  
 /protein\_id="BA02460.1"  
 /db\_xref="GI:398122"  
 /translation="MAEQMSRRTILAGALAGALAPVLATTSAWGGQAVRKATAAE  
 IALPRQVELVDPFVNAHSQVAGSPKVEETVYEEKIVIDAGTEVNAHAFNG  
 TVAPLMVIVHODVLETLINPTMTMNIIDPAATGALGGGLTEINRGETILRF  
 KATKPGFVYHCA.PGMAVPMHVVS.GMNNKIMVL.PREGIDHGKRLATYDKIVYVGGD  
 FYPRDENGKRYKYEAPGDVYEDTVYKMTLTPTTHYVFNCAVALGIDKMTAAVGEK  
 VLVHSQANRDRPHLIGHGIDVMAVGKNTPEVDQETWTFPGGAAGAAYTFPOOP  
 GIAYVYVNHNIIEAFELGAAAHFVTEGEMDDIMTSVLAPSGT"

BASE COUNT 540 a 661 c 663 g 487 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 1,48e-19 Length: 2351  
 Score: 427.50 Matches: 129  
 Percent Similarity: 47.00% Conservative: 51  
 Best Local Similarity: 33.68% Mismatches: 158  
 Query Match: 16.28% Indels: 45  
 DB: 1 Gaps: 14

US-10-088-045-2 (1-502) x AFANIR (1-2351)

Qy 15 LeuSerAlaLeuMetLeuSerGlyCysSerAsnGlnAlaAspLysAlaAlaGlnProLys 34  
 Db 787 TTGGCAGAGCCGCCCTGCTGGG-----GGCTTGGCGCGGT 825  
 Qy 35 SerSerThrValAspAla-----AlaAlaLysThrAlaAsnAlaAspAsnAla 50  
 Db 826 CTGGCAACACATCCGCTGGGACAAAGCGCGGTGCGGAAGCACTGCGGAGAAATA 885  
 Qy 51 AlaSerGlnGlnHisGlnGlyGluLeuProValIleAspAlaIleValThrHisAlaPro 70  
 Db 886 GCGACGCTTCACAGCCAG---AAGGTGAGACTTGTGACCTCTCGTGCAGGCCAT 942  
 Qy 71 GluValProProProValAlaAspAlaAspAlaLysValValValLysMetGluThr 90  
 Db 943 AGT-----CAGTTGCAAGAGCGGACCCAGAGTGTCAATTCACACATGTCATC 993  
 Qy 91 ValGluLysValMetArgLeuAlaAsp---GlyValGluTyrGlnPheThrPheGly 109  
 Db 994 GAGGAAAAAGAAATGTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1053  
 Qy 110 GlyGlnValProGlyGlnMetIleArgValArgGluGlyAspThrIleGluValGlnPhe 129  
 Db 1054 GGCACCGTACACAGACCGCTGATGTCGTCATCAGACATATTCGAACTGACACTC 1113  
 Qy 130 SerAsnHisPheAspSerLysMetProHisAsnValAspPheHisAlaIleThrGlyPro 149  
 Db 1114 ATCAACCTTGAAACCAACACGCTGATGACAAATATTCATGCGGACCGGATGCA 1173  
 Qy 150 GlyGlyGlyAlaGluAlaSerPheThrAlaProGlyHisThrSerThrPheSerPheLys 169

```

Db      1174 TTGGGGGGGGGGGGCTGACCGAATCAATCCGGGAGAAAAGACCAATCTCGCCCTTCAG 1233
      170 AAlaLeuGlnProGlyLeuThyValTyrHisCysAlaValAlaProValGlyMet----- 187
      1234 GCGACCAAGCCCGGCGCTCTTCGCTACACTGC-----GCACCTCCCGAATGGTCC 1287
      188 ---HisIleAlaAsnGlyMetTyrGlyLeuIleValGluProGlyGlyLeuPro 206
      1288 TGGCATGTGATATCGGCGCATGATGGATGCATGCTGCTCGCGCGAGGGCTTCGAT 1347
      207 -----LysValAspGlyGlyTyrTyrValMetGlnGlyAspPhe 219
      1348 GACGGCAAGCAAGCAACTGACCTACGACAAATATTATATCTCGCGCAACGAGATTTC 1407
      220 TTTT-----LysGlyLeuTyrGlyGlu-----GlnGlyLeuGlnPro 232
      1408 TATGTATCCGGCGACGACGAAAGGCAATACAGAAATACAGAGCGCGCGCGCTTAT 1467
      233 PheAspMetGlyLeuValAlaIleArgGlu---AspAlaGlyTyrValAlaPheAsnGlySer 251
      1468 GAAGACACCGGTCAAGGTCAATGGCGACTCTCACTCCGACCATGTGGTTCACAGCGCGCT 1527
      252 ValGlyAlaLeuThyGlyLeuGlnAlaLeuValAlaValGlyGlyLeuValArgLeu 271
      1528 GTGGGCGCACTGACTGCGCAAGAGCGCATGACGCGCGGTGGAG-----AAAGTC 1581
      272 PheValGlyAsnGlyLeuProAsnLeuThySerPheHisValIleGlyGluIlePhe 291
      1582 CTGATCTGCATCCGACCGCAAGCGCGCATACAGACCAATCTGATCGGGGGCGCATGGG 1641
      292 AspLeuValHisPheGlyGlyGlyLeuValAsnHisAsnIle-----GlnThrThr 309
      1642 GATTTATGTCTGGCGACCGCGCAAGTTCATACGCCCGCGACCTCATCAGAAACTGG 1701
      310 LeuIleProAlaGlyGlyAlaAlaIleThyGlnPheValAspValProGlyAspTyr 329
      1702 TTCATTTCCGGGTGTCGCCCGGACGAGCTTCTACACGTTCCAGACCGCGCATCTAC 1761
      330 ValLeuValAspHisAlaIlePheArgAlaPheAsnValGlyAlaLeuGlyIleLeuVal 349
      1762 GCGTATGTATACCAATCTGATCGAGCGTTTGAACCTCGCGCGCTCCGCCCATCTTCAAG 1821
      350 ValGlyGlyGlyLeuValAsnHisGlyLeuTyrSerHisLeuGlnThrAspAlaValTyrLeu 369
      1822 GTACAGGGGTGAATGGAACGATCTGATGACG-----TCGGTTCTCGCA 1866
      370 ProGlyGly 372
      1867 CCATCTGGC 1875

RESULT 32
AP005214/c 313200 bp DNA linear BCT 20-DEC-2002
LOCUS AP005214/c
DEFINITION Corynebacterium efficiens YS-314 DNA, complete genome, section 1/11.
ACCESSION AP005214 BA000035
VERSION AP005214.1 GI:23491835
KEYWORDS
SOURCE
ORGANISM
Corynebacterium efficiens YS-314
Corynebacterium efficiens YS-314
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacteriaceae; Corynebacterium.
REFERENCE
1 Kawarabayashi, Y., Yamazaki, J., Hino, Y., Kikuchi, H., Nakamura, Y.,
Ikeo, K., Suzuki, M., Mashima, J., Itoh, T., Yamagishi, A., Nishio, Y.,
Usuda, Y., and Sugimoto, S.
The entire genomic sequence of Corynebacterium efficiens YS-314
Published Only in Database (2002)
2 (bases 1 to 313200)
Kawarabayashi, Y., Yamazaki, J., Hino, Y., Kikuchi, H. and
Director-General of Biotechnology Center.
TITLE

```

## JOURNAL

## COMMENT

Submitted (17-MAY-2002) Director-General of Biotechnology Center, National Institute of Technology and Evaluation, Biotechnology Center; Nishihara 2-49-10, Shibuya-ku, Tokyo 151-0066, Japan (E-mail: bio@nte.go.jp, Tel: 81-3-33481-1933, Fax: 81-3-3481-8424) Kawarabayashi, Y. is officially affiliated with the National Institute of Advanced Industrial Science and Technology, Tsukuba, Ibaraki, 305-8566 Japan

Nakamura, Y., Ikeo, K., Suzuki, M. and Mashima, J. are at the National Institute of Genetics, Mishima, Shizuoka, 411-8540 Japan Itoh, T. is at the Japan Biological Information Research Center, Koto-ku, Tokyo, 135-0064 Japan

Yamagishi, A. is at Tokyo University of Pharmacy and Life Science, Hachioji, Tokyo, 192-0392 Japan

Nishio, Y., Usuda, Y. and Sugimoto, S. are at the Ajinomoto Co., Inc., Kawasaki, Kanagawa, 210-8681 Japan

The other authors are at the National Institute of Technology and Evaluation, Shibuya-ku, Tokyo, 151-0066 Japan.

## FEATURES

## source

## CDS

```

1..313200
/organism="Corynebacterium efficiens YS-314"
/mol_type="genomic DNA"
/strain="YS-314"
/db_xref="taxon:196164"
1..1722
/note="CE0001, similar to AF222789-7|AAF33692.1| percent
identity: 48 in 567 aa"
/codon_start=1
/transl_table=1
/product="putative chromosomal replication initiator
protein DnaA"
/protein_id="BAC16811.1"
/db_xref="GI:23491836"
/tranlation="MSONSSSLLETWROVENDLNLSSQODSGIPLTPTORAYLTLA
KRIALVDGAYLSTHALAKNAIENDLESRLTKLSMRGRFSIAYSEPRDDEP
NAPAPRQDELNYPYFGQPOSQGGQCGQHAPVOEVRAPAPHQGGQQAQHQ
PRANAPQOYVVGSGRQASQAGAMEDTHSPAPDOGDPAPVEPDPAPHPLRI
PRETPAHNPRESLNPRTYFENFVIGFNPANAAVAAPASPAKAFPLFISGGG
LGKTHLAVNGAOELQGLRIKYVSESEFPNDYINSVRODROSPKRYRNILDM
VDDIOFLAGKEGTQSEPRFPALHQAOKIILSSDPROCLTIEDRLRTPREGLI
TDIOPDLETRIALIMKKAQADGTHVDEBELLSREBSIRELEGALIRVASSLI
INQPLDKEMALVALDILPEPDMETITAPVIMEVAEYFKISVDTLRGAKYRAAHA
ROLAMYLCRELIDSLPRKIDGVFGKDHITWYVADRKIROEMTERDRTDEIQOLTL
IKSRGRN"
complement(2289..2603)
/note="CE0002"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="BAC16812.1"
/db_xref="GI:23491837"
/tranlation="WMUHADIKLAEKVIKCRTRGCHDPTOPDPADTVVWAGRL
PWTRAIGVVELHLPNVGHLVAREFVIFLCRSERCPEARFLSFTNVSPPHPLLV
KF"
2592..3776
/note="CE0003, similar to L39923-25|AAB53142.1| percent
identity: 50 in 390 aa"
/codon_start=1
/transl_table=1
/product="putative DNA polymerase III beta subunit"
/protein_id="BAC16813.1"
/db_xref="GI:23491838"
/tranlation="MEPONVSFVRAVREDLTVSGVAVANLPTKTPQVIRLMTAAD
EGLEAGDVEYSRVRSAVAYVGRALAVAKKILTEVITIGTPKRPVPRLDGKLV
SGSASFELPSIPLDYPPLPSPERTSINPKLATEVTOVASAGBDSDLPMLTGV
MEIHGNGQISLAATDFRLARTEFWDASGIEKKLIIPAKTLLENRTLDGTANE
IGIIVGTQVQVREGFLGIRTDNRESTRMADAPNLIPLPKHETLIVASVEIANPL
DAVRVSLVADRNQOIIIEFSDGVITLSAGSDAGHAQVPLPCATGDKLKAFFGY
LKDGSLVPTSRVAVGFEPERPAPMIPEPEPEPADRGVFGPTDVTFTYLLMVRLL
PC"
3878..5074
/note="CE0004, similar to AF222789-9|AAF33693.1| percent
identity: 55 in 390 aa"
/codon_start=1

```

/transl\_table=1  
/product="putative DNA replication and repair protein  
Ref"  
/protein\_id="BAC16814.1"  
/db\_xref="GI:23491839"  
/translation="MYTRSLRLDFRFRPELKYELKGTITFIIRNGCFKTNIVEAIG  
YLHLSSHRVSTDAFLPRAVAGDARISAVNQGRELAAHLIKPHANQGIINIRKV  
KSPRELKIVTLFAPEDLAVRGEPAERRYLDIATRPRMVKADYDKVLR  
RNALKTATIALRRYGTIEGAALALITDMDGLARLGAEMVAFALVQDLSQIR  
DAYOTIAEERPAVNYKTTIDQAGFEFDGATITELTAAKROREIRGMSL  
VGPBHDLELHIGOPAKGFSGETSPALSMRIAEFNLIRSDGDPILLIDVSE  
LDAGRREKLVGIARDAEQVITTAASDDEPAMADAITAHTVTVPDITDAGRISLDE  
QDPDP"  
5071..5652  
/note="CE0005, similar to AF222789-10|AAF33696.1| percent  
identity: 39 in 188 aa"  
/codon\_start=1  
/transl\_table=1  
/product="conserved hypothetical protein"  
/protein\_id="BAC16815.1"  
/db\_xref="GI:23491840"  
/translation="MTEDDFIAQAFERVRRAQRNRGRVPLDNNRPAFRRTPRIG  
DRANGGGGGVDPRRRGPTGDCRYKRYSAEPLGAILNRIRSGMSDIAAG  
WTSHWELVGPKIAGTKVEMTKDKLFTICSTAWATNLRMQKQILQYIAEKVG  
DVIAELRIFGPQAPSWRKGPLHVKGPRDTYG"  
5786..7849  
/note="CE0006, similar to X92503-7|CAA63253.1| percent  
identity: 72 in 685 aa"  
/codon\_start=1  
/transl\_table=1  
/product="putative DNA gyrase subunit B"  
/protein\_id="BAC16816.1"  
/db\_xref="GI:23491841"  
/translation="MANTENYDASITILEGLAAYRKRPMYIGTSGRLHLIME  
VUNSDAMAGYADRVDTLLLEDGVEVINDNGRGIPEMHAAGAPVYVMOLAG  
GKEDSDSYAVSGELHGVISVNALSTRVDAHKKQCKHYQNPNSIPEDVLEGAN  
RGITGIRFWDALIEFTDPTDITSRQEMAFNLKGLTWLTDNRASDELEA  
LAEGDTAVELSDQIDADETQLVDAEVDAPKPKRKKIKFHYNGLEBYVNL  
NRTKPIHPSIVSFEAKAEVEIAMQMSYKESVHTANTINTEGTHGEFSS  
ALTSUMRWYRHKHLKEKANLTGDDCGSLSVISVRSEPOEQTQXKGNIEI  
KSPVQWANHIGHMLEANPAKVIINKVAGSDNRQATLPRLGKLYNERKLDV  
PKLADCRSKDPESELYIEGDSAGSAGSDSMQATLPRLGKLYNERKLDV  
LKSEVQAITTALGTGIDHDFINKLRYHKIYLMADADVQGHATLLTLFRFMD  
LIAEGHYLLAQPLVYKLMQGEFGFAYSDSEBDQOLEGALGRKINKDDGIQRYG  
LGEMLPMLMETTMDPSVIRIRVDLADAGRADELFIILGMDVVARSPITNADV  
RPLDV"  
7945..8847  
/note="CE0007"  
/codon\_start=1  
/transl\_table=1  
/product="hypothetical protein"  
/protein\_id="BAC16817.1"  
/db\_xref="GI:23491842"  
/translation="WNAETPTTVNHEMPSSRTPVQVVRAGEDQSGSGDKLVNWM  
PGMGGYFRPIAELARGFPAVAGELGQGSRAATRTTWGHDLASDFPQ  
VRAKOGLGPDHPMIFLAHSGQISCLFARBEDRDNLGIFGVGSGSYRPF  
DPMGRILGWLGLFGRVGGHGLGYMKGKILGRDPLGFGAAGAHMERLPRKNSL  
DILTGDIDVVALSRLELPLVLFRTNRDNCPTASEALAEFLHNDVAFEDLPGL  
GHRMAEPIYVADRFVSPAGVLD"  
9128..9709  
/note="CE0008"  
/codon\_start=1  
/transl\_table=1  
/product="hypothetical protein"  
/protein\_id="BAC16818.1"  
/db\_xref="GI:23491843"  
/translation="MAEKVHVPVGEHAHHEAGVNAVPSAVERTVDDHSDAQA  
WICTDHPVANNRINISGAIAGAVTIALVLVIGIGAAAGLQDSSATVIGITVLQ  
VIAFLVAGYVAGLAVRGLHGLFGLCATGSLAVIFLAGWLGSIVLGAPEGIAGTAD  
TAAOTVNTSODQGAADATSGEDVDQAQAEAE"  
9994..10422  
/note="CE0009, similar to X65467-3|CAA64662.1| percent  
identity: 35 in 113 aa"

/codon\_start=1  
/transl\_table=1  
/product="putative transcription regulator"  
/protein\_id="BAC16819.1"  
/db\_xref="GI:23491844"  
/translation="MAGVRKQNPRTIGQLVYGOHMLTASRLDVMNRIGRAMDPT  
SRLSLDAPGPAQATLGLTIRTNVSNHCLDCCGIVAAEPGRRTREIADAH  
LRAALNALDITLVAIDENADCIDPQPVNCCAPBEGERSQ"  
10419..12371  
/note="CE0010, similar to 274025-13|CAA98411.1| percent  
identity: 55 in 627 aa"  
/codon\_start=1  
/transl\_table=1  
/product="putative cation-transporting ATPase"  
/protein\_id="BAC16820.1"  
/db\_xref="GI:23491845"  
/translation="MTSACGCEPTGCVDPADPVMAAVGTEHPTPMWRDPTVPIV  
ASGVALSGILIEWFVPGAGTASVFWISLIGASQFPGAVRNILVTRGLIGLIM  
TTSATGAVILGFVEEAMALAFYSIALEAGRAMDAPRGLRLAIPATVAVTTA  
NTRSVSEELLPGDRVLVAGERLATDGIHSGRSLDVSATGESIPEVGPQOV  
LGSINISGLVEEATPAAGTDSLTVVLVEQAQHEKDRADRLARLPVGVLI  
LALVAVISLIGDPVWTERLVLVVAASPCALISVPTVSAIQAASRPFVIXS  
GAAFELSGIOHLADKCTIRNOPTVADVVTGASPSRAVLGMAALAEASTHPLA  
AATTAATPEVPAAEYEVAGSGITGLVGTIRIAGVSPMLAPGLVGTVEDEGKM  
TVVAVIRDOQPGVIGVDELREPDPVIRLETQGFITMLDQRTAEALATAG  
IGDVHAQLRPRDKARAVAGIAGRGVAMIGDINDAPALASADVIGAMGKDAVE  
SADVARTGDDLRVTPALHARRGRTINQNIIVLSAIIIVLPLAITGALGAAYVL  
VHEVAEVIYIINGLRAARAGNPALS"  
complement(12368..13030)  
/note="CE0011"  
/codon\_start=1  
/transl\_table=1  
/product="hypothetical protein"  
/protein\_id="BAC16821.1"

CDS

## Alignment Scores:

Pred. No.:	1,15e-14	Length:	313200
Score:	390.50	Matches:	108
Percent Similarity:	45.06%	Conservative:	47
Best Local Similarity:	31.40%	Mismatches:	130
Query Match:	14.87%	Indels:	59
DB:	1	Gaps:	13

US-10-088-045-2 (1-502) x AP005224 (1-313200)

Qy	31	AlaGlnProLysSerSerThrValaAlaAlaAlaThyThraAlaAsnAla	50
Db	298126	GCCGCCCGGGCGGGAACCAACCGCGCGGTGACGTACCCGAGTC---GGCGCC	298070
Qy	51	AlaSerGlnGluHisGlnGlyGluLeuProValIleAspAlaIleValThraAlaPro	70
Db	298069	CGCAGCGTGACACGACGAGGCGTTCACCCCGGTCAACCCGGTCTGCTG-----	298025
Qy	71	GluValProProProValaAspArgAspHisProAlaLysValValaLysMetGluThr	90
Db	298024	---GCCCGCGCGCGCGCGGTACCGTCCACGAA-----TACAGCTGGGATATC	297980
Qy	91	ValGluLysValaMetArgLeuAlaAspGlyValGluTyxGlnPheTrpPheGlyGly	110
Db	297979	ACCGAGAGGTCCGCCAGGTCGACCGGGTCATGACAGGTGCTGCGCTTTCGACGGG	297920
Qy	111	GlnValProGlyGlnMetIleArgValaArgGluGlyAspThrIleGluValGlnPheSer	130
Db	297919	CAGCGCCCGGACCGACCGCTCGCGGCAAGCTCGGGACACCTTCGCAATCACCCG--	297863
Qy	131	AsnHisProAspSerLysMetProHisAsnValaAspPheHisAlaAla---ThrGlyPro	149
Db	297862	---CAACAACAGGCGACATGATCATCTCATGATTTCCACCCCGCGAGGTCAACCCG	297806
Qy	150	GlyGlyGlyAlaGluAlaSerPheThrAlaProGlyHisThrSerThrPheSerPheLys	169
Db	297805	-----GACACCAACATGGCGACACCCCGTCCGGGGAATCCTTGACATGAGATTCGTC	297752



gene  
3389. .4528  
/gene="pan1"  
/note="synonym: VNG1187G"  
3389. .4528  
/gene="pan1"  
/note="pan1"  
/codon\_start=1  
/transl\_table=1  
/product="membrane protein"  
/protein\_id="AAG19561.1"  
/db\_xref="GI:10580721"  
/translation="MTDQIGAPLGISRRDVAATAGVGTAAIAGCTAPDGGSEVTTT  
TIAAKQSGLPPTSPPEVDATGQGNQVTLKSPAVADVPLDSMGKGVKLERVAFAFA  
EGDGPVSPPIVTEBEGDIEVTLNDTGGKPHILHFHGSQTAWEDDVPPTTGI RVG  
PEELKTYTIPANVPGTHLYHCHYQTRHIDMKMYGIPRIDKGYEPADKEFMTVKDM  
BSRLNYSMTAGEVDVSPRTNRPNDFVTKGSAPEFLHEDGSPILIVEGDIVRHLIKM  
GGMNRPRLHINHRFRMVEKDGQIPAAAGHMDITDAPAPERTIEQADAPGIVL  
MCHKKNHVMGTFYFGMLSGIVTKMDNDISFGLKMDPAGYDA"  
4535. .5152  
/gene="hcpD"  
/note="synonym: VNG1188G"  
4535. .5152  
/gene="hcpD"  
/note="hcpD"  
/codon\_start=1  
/transl\_table=1  
/product="halocyanin precursor-like"  
/protein\_id="AAG19562.1"  
/db\_xref="GI:10580722"  
/translation="MTDSDAVTRRRVIGSAGAGAAAGITGFPAGGAQSASIDDM  
LDVSNYADONREDTSSSEVTVDVGDANGAVGEPAAVRVDPTVTFFENVSNTNHN  
AVESQPSDMDNGVGEIINNEGTSHTSETAGVYFCSPHRANGMKCAVVGADVGV  
GAAGSIDMGNGVIGAGFALGIGLMPFVRALDSPHNNTHE"  
complement (5173. .5571)  
/gene="VNG1189H"  
complement (5173. .5571)  
/gene="VNG1189H"  
/note="hypothetical protein"  
/codon\_start=1  
/transl\_table=1  
/product="Vng1189h"  
/protein\_id="AAG19563.1"  
/db\_xref="GI:10580723"  
/translation="MTAARVNPPTVMVRLRGDQHPRLVLVGDSTATVIERHV  
ATLVAGGDTGDAITLALRGITDGARRAALANGVANVAATIRNDVPTTHILIV  
ATADDAKRSVPAATTDDSAVAADATDT"  
complement (5721. .6323)  
/gene="sod1"  
/note="synonym: VNG1190G"  
complement (5721. .6323)  
/gene="sod1"  
/note="sod1"  
/codon\_start=1  
/transl\_table=1  
/product="superoxide dismutase"  
/protein\_id="AAG19564.1"  
/db\_xref="GI:10580724"  
/translation="MSEYELPPLPYDIYALERPHISEOVLTHHDTTHQGVYNGKDAE  
ETLVANRETDHSAAGLADGVTHNGSHIILHTLFQWMSFAGDEBSGLAADIAD  
FSSGNMRAEPFAAASAGMALVYDHSHTLRNVAVDNDEGALWGSHPIALDW  
EHSYVDYDQDGSFVDAFFEVVDWDEPTERFEQAARFE"  
6585. .7607  
/gene="acd3"  
/note="synonym: VNG1191G"  
6585. .7607  
/gene="acd3"  
/note="acd3"  
/codon\_start=1  
/transl\_table=1  
/product="acyl-CoA dehydrogenase"  
/protein\_id="AAG19565.1"  
/db\_xref="GI:10580725"  
/translation="MSEYELPPLPYDIYALERPHISEOVLTHHDTTHQGVYNGKDAE  
ETLVANRETDHSAAGLADGVTHNGSHIILHTLFQWMSFAGDEBSGLAADIAD  
FSSGNMRAEPFAAASAGMALVYDHSHTLRNVAVDNDEGALWGSHPIALDW  
EHSYVDYDQDGSFVDAFFEVVDWDEPTERFEQAARFE"  
6585. .7607  
/gene="acd3"  
/note="synonym: VNG1191G"  
6585. .7607  
/gene="acd3"  
/note="acd3"  
/codon\_start=1  
/transl\_table=1  
/product="acyl-CoA dehydrogenase"  
/protein\_id="AAG19565.1"  
/db\_xref="GI:10580725"  
/translation="MSEYELPPLPYDIYALERPHISEOVLTHHDTTHQGVYNGKDAE  
ETLVANRETDHSAAGLADGVTHNGSHIILHTLFQWMSFAGDEBSGLAADIAD  
FSSGNMRAEPFAAASAGMALVYDHSHTLRNVAVDNDEGALWGSHPIALDW  
EHSYVDYDQDGSFVDAFFEVVDWDEPTERFEQAARFE"

[illegible]



QY	151	lyglalaglaalaserPhe-----Thra	159
Db	3809	--GGATCAGACAGCGCTTGGGAGAGACGCGCTCCCGACGACCGCGCATCCGGGTGC	3866
QY	159	laProglyH1srHserThePheSerPheHylsalauInProglyLeuYrValYrH	179
Db	3867	GGCCCGGGGAGAAACACCTACACCTTCCGCCAACGTCGGGAAACGACCTCTACC	3926
QY	179	isCyAlaValAlaProValIglyMechisIleAlaInglyMeTyrgIyleuIleuV	199
Db	3927	ACTGCCACTAC---CAGACCCAGCGCCACATTCACATGGGAGATGTACGGCATCTTCCGA	3983
QY	199	alGIuProLySGluGIuLeuProLyValAspLyGIuYrValYrValMetGIuAspP	219
Db	3984	TCGACCCCAAG---GGGTACGAGCCCGCCGACAAGAGATATTCATGACGCTCAAGACT	4040
QY	219	heYrThrYrSGlyYrYrYrGIuGIuInglyInProPheAspMetGIuYrAlaI	239
Db	441	GGGACTCGCGCGTGAACCGCAGCATGGCCGCGC-----	4072
QY	239	leargIuAspAlaGIuYrVal-----ValPheAsnGIuYrValGIuAlaLeuThrg	257
Db	4073	-----GAGGACGTGACTACAGCCCGCGCACCCGCAACCCGCGTTCACCGTCAACG	4127
QY	257	lyGIuAsnAla-----LeuYrAlaYrValG	266
Db	4128	GCAAGTCCGCGCGCGAACCTCTCACCCCGAGAGACGGCTCACCGCATCTGCGAGCAGG	4187
QY	266	lyGIuThrValArgLeuPheValGIuAsnGIuYrProAsnLeuThrSerSerPheH1sv	286
Db	4188	GTAATACGCTCGCGCTCACCTGCTACAGCGCGG---TACATGAACACCCCGCTGCACA	4244
QY	286	alIleGIuIlePheAspLyValH1sPheGIuGIuYrYr-----GIuGIuA	303
Db	4245	TCGACACACACCGCTTCGGCATGTGCA7AAGAGACGGCGGCGGAGATCCGGCGCGGCC	4304
QY	303	snH1sAnIleGIu---ThrThreulIleProAlaGIuYrAlaIleThrGIuPheI	322
Db	4305	AACACACACATGACATCACGACATGGCTCCCGCCGAGCGCACACCATC---GAGTTCC	4361
QY	322	ysValaAspVal---ProGIuAspYrValLeu-----ValAspH1sAlaIleP	337
Db	4362	AGCGCGGACCGCGACCCCGGATCTACCTGATGCACTGCCACAGCAAGACGATTATGA	4421
QY	337	heargAlaPheAsnLySGIuAlaLeu	345
Db	4422	ACGCGACGTTTACCCCGGTGCATG	4447
RESULT 34			
LOCUS	AE011334	10485 bp	DNA linear BCT 28-APR-2003
DEFINITION	Leptospira interrogans serovar lai str. 56601 chromosome I, section 143 of 397 of the complete sequence.		
ACCESSION	AE011334	AE010300	
VERSION	AE011334.1	GI:24195164	
SOURCE	Leptospira interrogans serovar lai str. 56601		
ORGANISM	Bacteria; Spirochaetes; Spirochaetales; Leptospiraceae; Leptospira.		
REFERENCE	1 (bases 1 to 10485)		
AUTHORS	Ren, S. X., Fu, G., Jiang, X. G., Zeng, R., Miao, Y. G., Xu, H., Zhang, Y. X., Xiong, H., Lu, G., Lu, L. F., Jiang, H. O., Jia, J., Tu, Y. F., Jiang, J. X., Gu, W. Y., Zhang, Y. Q., Cai, Z., Sheng, H. H., Yin, H. F., Zhang, Y., Zhu, G. F., Man, M., Huang, H. L., Qian, Z., Wang, S. Y., Yao, Z. J., Shen, Y., Qiang, B. Q., Xia, Q. C., Guo, X. K., Danchin, A., Saint Girons, I., Somerville, R. L., Wen, Y. M., Shi, M. H., Chen, Z., Xu, J. G. and Zhao, G. P.		
TITLE	Unique physiological and pathogenic features of Leptospira interrogans revealed by whole-genome sequencing		
JOURNAL	Nature 422 (6934), 888-893 (2003)		
REFERENCE	2 (bases 1 to 10485)		
AUTHORS	Ren, S., Fu, G., Jiang, X., Zeng, R., Lu, L., Lu, G., Jiang, H., Ding, Y., Jia, J., Tu, Y., Gu, W., Cai, Z., Sheng, H., Yin, H., Zhang, Y., Zhu, G., Wang, S., Shen, Y., Qiang, B., Chen, Z., Wen, Y., Xu, J. and Zhao, G.		
TITLE	Direct Submission		
JOURNAL	Submitted (12-MAR-2002) Chinese National Human Genome Center at Shanghai, 250 Bi Bo Road, Shanghai 201203, China		
COMMENT	Updated information will be available at our World Wide Web site ( <a href="http://www.chgc.sh.cn/lep/">http://www.chgc.sh.cn/lep/</a> ). Comments to the authors are appreciated.		
FEATURES			
source	Location/Qualifiers		
	1..10485		
	/organism="Leptospira interrogans serovar lai str. 56601"		
	/mol_type="genomic DNA"		
	/strain="56601"		
	/serovar="lai"		
	/db_xref="taxon:189518"		
	/chromosome="I"		
	/note="Strain 56601 is maintained by the Institute of Epidemiology and Microbiology, Chinese Academy of Preventive Medicine"		
gene	complement (90..1391)		
CDS	/locus_tag="LA1552"		
	/codon_start=1		
	/transl_table=11		
	/product="Sensory transduction histidine kinases"		
	/protein_id="AA048751.1"		
	/db_xref="GI:24195165"		
	/translation="MLPRNRVNVWIDMNLSPFKLFLSHRSYLSGVKKNVIO IILIVSCRGFLIYRPSYLAWOGGSGVVLDTLALLVFLFLLRNRPKPSYTL SVLIFILGGLTIGLGLGAGILMLFVPPVCGIFLGRILITFMSVITSLVESGILF HVKITWTGITSFPQVVGANFTFLCGITFLPILIEINBAKQSHLLTHNTNKKIQ NKIPAYTEKESGILBSLNYKETFLKEIDIRKVNHHILFLNDRKASELAKTII ESAIDRIOSMGIVQVEYFLANDYSKYVIOKEVLDLTNLNHLFLSYRQSGISOVNINTE ALISYDKAIPCGILINELINMINSFGHAPENRKIKTIVSFQSDSGYFLIEADNGIQ KSEKSKMAGMSLIAFLQJRGIEIIEKENGFLVRVRFPP"		
gene	1635..3656		
CDS	/locus_tag="LA1553"		
	/locus_tag="3656"		
	/locus_tag="LA1553"		
	/codon_start=1		
	/transl_table=11		
	/product="putative lipoprotein"		
	/protein_id="AA048752.1"		
	/db_xref="GI:24195166"		
	/translation="WKILNPEITVVVYVFWLMTGISRGIFISCSGENSENKEISGFH VLTINDSFPDMOKIPVGGIYEFINSGNPNNAIADKNMSTKSPGNIYMSRTKTK VSPKRGVPPYFCSFHATPDGKNMVGDIIVGANVPYPAASGSMKNVDSFSTGRK VPSQYPTIONAVDASNPDLVILSEGILYELBEVATVTSITIRGDRKVIIDGPFQSG NGIMVAVDGVVIEMLTARNATLNGFYWGKRGSLTATHNNGDIYAFDSINCV IESHVASGSDSGIYIGOCYPCAKIIVDVVEHNAIGSGNAGELYLIGSVKNNI VGLANTDIRELIPPERETITLGNLVYNNNPKAPIALAEYPSGNGIILAGLCSVIG RKQVVIENONGIVILPNLDENFMLSNNIYODNIIVNSGADITLLGPMSTGCSNG NNEYRTLPFLKMGCGSWIRLPVGGDLSMGLALGLMVASGSRPSGNYEQPIP KQLMNPGLNAPVKEPALTAFFEDFNILNINQVLEPKRAEILIKTVPRAPASTGATILV KGIQGFPEFYHMLGLPFAIYICWTSMSLIDLRDLEWIRIKIYIVITLIVPILIS PAIYLIVGGVGLALILIVTSVAKP"		
gene	3631..4011		
CDS	/locus_tag="LA1554"		
	/locus_tag="4011"		
	/locus_tag="LA1554"		
	/codon_start=1		
	/transl_table=11		
	/product="hypothetical protein"		
	/protein_id="AA048753.1"		
	/db_xref="GI:24195167"		
	/translation="MLBPKRSVYKFKIIRNYRRKIMEQIVWVGPGFALLPFPYGYR PFIYLYTLALPLALSDIVKREDVNSKISWTGAILIPILAGAYLVWAGSKIPSWL KNIIVYGGVGLALILIVTSVAKP"		
gene	3998..5029		
	/locus_tag="LA1555"		

gene 3998. .5029  
CDS /locus\_tag="LA1555"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AA048754.1"  
/db\_xref="GI:24195168"  
/translation="MNLPRKKLNKQPLKWTIGISAGATGAGTSSIDEDSGLLCKTIR  
RITLVNNEKRTIPGNPNNSYGSWVHPPEFHIDPTFLNRMELKNLEISSGPILOKHS  
IYEMPTLVANHTVAVKMTNGVLPGPVPVRAKLGQMEITLRNDSHPHSIHFHSDHP  
NEBGMFLVVAGSESKTYQLTAPGIGFHPYVCHVPPLASHMAGLVXGGLIVDPGGRPPA  
HEFMFLSMQDLEDJEGKNDLFCMNGMAGFYDYPPIKVPVGOKVRLXIYNMCEYEPVAS  
FHLHAQTFPVFPTGKTLIPDHDVTVTIGQTERVILERTLPKRGYMHHPQTMAEK  
GAMGWIVAA"

gene 5376. .5369  
CDS /locus\_tag="LA1556"  
5376. .5369  
/locus\_tag="LA1556"  
/codon\_start=1  
/transl\_table=11  
/product="SCOl/Senc family protein"  
/protein\_id="AA048755.1"  
/db\_xref="GI:24195169"  
/translation="MKERILLISILIGILGFGPMKAKGRSSIOEPVFEVSEQITIT  
LNTNOGLGPIUNKIPGKTLVYRGSHCPDMCPRLIDMSLAVKEIGOGKNTLPVET  
SYDPRDPTSLAKTYKQFPQERLVALTGEKSKDLSLONAGAVSKVDAPQLGCGYTT  
VHVHYIVLVDKRSRLMTFPGMDGKTLAKEIRRL"

gene 6040. .6186  
CDS /locus\_tag="LA1557"  
6040. .6186  
/locus\_tag="LA1557"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AA048756.1"  
/db\_xref="GI:24195170"  
/translation="WQPEKILHEHSIDELKNSIIDSKIVSMHPRATNTNELLF  
QQLY"

gene 6229. .6345  
CDS /locus\_tag="LA1558"  
6229. .6345  
/locus\_tag="LA1558"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AA048757.1"  
/db\_xref="GI:24195171"  
/translation="MILPEIKKKIQDRLVEIESRSVEILLAISSGRAWCF"

gene 6314. .6529  
CDS /locus\_tag="LA1559"  
6314. .6529  
/locus\_tag="LA1559"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AA048758.1"  
/db\_xref="GI:24195172"  
/translation="MNQVQLGFPESIDSDCDIRFIYGHETEMFLAVLPKKMSLFPYK  
INLVRIFPCARDRLGVKLLIRPSKWL"

gene 6747. .7055  
CDS /locus\_tag="LA1560"  
6747. .7055  
/locus\_tag="LA1560"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AA048759.1"  
/db\_xref="GI:24195173"  
/translation="MTFAMVGHILKEFFHQSPLPTIFLSWGIALFEYILMVPANRIGYN  
BEGYSTPQLKIQEIIITISVFIPLFASLVLKERIKMNHAVSFLILAAVGFAPDKTSS  
"

gene 7063. .7176  
CDS /locus\_tag="LA1561"

CDS

7063..7176  
/locus\_tag="LA1561"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AA048760.1"  
/db\_xref="GI:24195174"  
/translation="MEHISRKNVFIEKEYEVVSHKGNSDLIIVNLFCTL"  
7632..7736  
/locus\_tag="LA1562"  
7632..7736  
/locus\_tag="LA1562"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AA048761.1"  
/db\_xref="GI:24195175"  
/translation="MROWLRRLTLPIKIFRLDLDLIPKGFSPFKGT"  
complement(8434..8826)  
/locus\_tag="LA1563"  
complement(8434..8826)  
/locus\_tag="LA1563"  
/note="confirmed by proteomics"  
/codon\_start=1  
/transl\_table=11  
/product="class II heat shock protein (HSP20)"  
/protein\_id="AA048762.1"  
/db\_xref="GI:24195176"  
/translation="MTNAVINERKHSVSEETKEVRILAPVDIYSDENIYLLAD  
LPGEKRVQVQLEKDOIISGKTSNKDIOGELNISEPRTGEYKRTITLVESVEDRI  
SAVYKNGVLTILPKRKPILKIKIVRS"  
complement(8847..9263)

gene

CDS

gene

Alignment Scores:			
Score:	1.71e-07	Length:	10485
Posed:	258.00	Matches:	78
Percent Similarity:	42.96%	Conservative:	41
Best Local Similarity:	28.16%	Mismatches:	150
Query Match:	9.82%	Indels:	38
DB:	1	Gaps:	11
US-10-088-045-2 (1-502) x AEO11334 (1-10485)			
QY	75	ProValaAspArgSPHISProAlaIysValValValLysMetGluThrValGluIysVal	94
		:::     :::	
DB	4280	CCGATCTTAAACAACAT-----ATTAGCTTTAGAAAACCT	431
QY	95	MetArgLeuAlaAspGlyValGluIyrGlnPheTrpThrPheGlyGlyGlnValProGly	114
		:::     :::	
DB	4319	TTGACGGTTGGGCACAAATACAGTACTCAAGCTTGACCTTTTAACGGTTTAGTCTGGT	437
QY	115	GlnMetIleArgValArgGluGlyAspThrIleGluValGlnPheSerAsnHisProAsp	134
		:::     :::	
DB	4379	CCTGGGTTCCAGACCAATCGGACAAAGATGAAATCATCTTAGAAAC-----GAT	443
QY	135	SerIysMetProHisAsnValAspPheHisAlaIleThrGlyProGly---GlyGlyAla	153
		:::     :::	
DB	4433	TCTGAAACATCCACATTCGATTCACCTTACACGGAAATCATGATCCGAAACGAGCGTTGG	449
QY	154	GluAlaSerPheThrAlaProGlyIHisThrSerThrPheSerPheIysAlaLeuGlnPro	173
		:::	
DB	4493	GAGCCG-----GTAAGCGTTGGTAGTGA AAAAGGATCAACTGACCTGGGTCGCAAT	454
QY	174	GlyLeuIyrValIYTHIECyAlaValAlaProValIglIyMetHisIleAlaAsnGlyMet	193
		:::     :::	
DB	4547	GGATTTCATCCGATCATCTGTCACTTCTCCTCATTTGGCGTAGTCATATGCGCAAAAGCTTT	460
QY	194	TyrGlyLeuIleLeuValGlnProIysGluGlyLeuProLysValAspIysGluIyrTyr	213
		:::     :::	
DB	4607	TACGGAGGTGTGATCGATGATCCACTGAGAGTCTGCGCCGTCGACATGAATTTATAGCTG	466
QY	214	ValMetGlnGly---AspPheTyrThrIysGlyIysTyrGlyGluGlnGlyLeuGlnPro	232
		:::     :::	

```

Db      4667 ATTCTTTCGATGGATCTAGAACAAAGGAAAG----- 4702
Qy      233 pheasmetclulysalaileargluspalaiglutryvalphasniglyserval 252
Db      4703 -----AATGATCTTTTGTGGAAACGGAATGGCT 4732
Qy      253 glyalaleuthrclglusnalauleulysalailevalglutthralphleuphe 272
Db      4733 GGT---TTTACGATCGTATTCGATTAAGTTCCAGTACGTCMAAGGACCTCTTAT 4789
Qy      273 valglasniglyglypro---asleuthrserSerPhehisValIlelglulieppe 291
Db      4790 ATCCGAAACATGCTGTAATGACCCGTTGCTTCCTTCATCTACACGCTGACATTC 4849
Qy      292 asplysvalhisPhehiglulglulysglulysnhisasnileglinthrthleulle 311
Db      4850 GACGTC-----TTTGAACCGGAGCTAATTGATTCCTGACATCATACGACGTAGTT 4903
Qy      312 ProAlaGlyValA-----AlaIlethrGluPheLysValAspValProGlyAspTyr 329
Db      4904 ACTCTGGGGCAACACGAAAGGTGATTTTGAATTCACTTCACAAAGAGGAGGTAT 4963
Qy      330 ValIeuValAspHisAlaIlePheArgAlaPheAsnLysGlyAlaIleuGly 346
Db      4964 ATGTTTCATCTCAACCAACAAATATGCG---GAGAAAGGGGCTATGGGT 5011

RESULT 35
AF339049 537 bp DNA linear BCT 15-MAY-2001
LOCUS Nitrosomonas sp. TA-9211-NH4 putative dissimilatory nitrite
DEFINITION reductase (nitrK) gene, partial cds.
ACCESSION AF339049
VERSION AF339049.1 GI:14039642
KEYWORDS
SOURCE
ORGANISM Nitrosomonas sp. TA-9211-NH4
            Nitrosomonas sp. TA-9211-NH4
            Bacteria; Proteobacteria; Betaproteobacteria; Nitrosomonadales;
            Nitrosomonadaceae; Nitrosomonas.
REFERENCE 1 (bases 1 to 537)
AUTHORS Casciotti, K.L. and Ward, B.B.
TITLE Dissimilatory nitrite reductase genes from autotrophic
JOURNAL ammonia-oxidizing bacteria
MEDLINE Appl. Environ. Microbiol. 67 (5), 2213-2221 (2001)
PUBMED 11319103
REFERENCE 2 (bases 1 to 537)
AUTHORS Casciotti, K.L. and Ward, B.B.
TITLE Direct Submission
JOURNAL Submitted (22-JAN-2001) Geosciences, Princeton University, Guyot
            Hall, Princeton, NJ 08544, USA
FEATURES
source 1.537
            location/Qualifiers
            /organism="Nitrosomonas sp. TA-9211-NH4"
            /mol_type="genomic DNA"
            /strain="TA-9211-NH4"
            /db_xref="taxon:159355"
            <1..>537
            /gene="nitrK"
            <1..>537
            /gene="nitrK"
            /codon_start=2
            /transl_table=11
            /product="putative dissimilatory nitrite reductase"
            /protein_id="AAK53363.1"
            /db_xref="GI:14039643"
            /translation="VYHCAPEGMVPMHVVSGNHGTTIMVLPNGLKNEPKPLHYDRY
            YIGENDFYIPKDENGKYKTYDPSGEAFDTTIDMRLLIPNHVFNRGVALTGKAMT
            AKVGEVMEVHSEANDRTPHILIGLGDVMEYTGKSNPEKDLFTWFIKSGSAGAL
            YTFRQGVAYVNNHLEA"
gene
CDS
BASE COUNT 122 a 165 c 151 g 98 t 1 others
ORIGIN
Alignment Scores:

```

```

Pred. No.: 2,32e-08 Length: 537
Score: 249.00 Matches: 73
Percent Similarity: 52.20% Conservative: 22
Best Local Similarity: 40.11% Mismatches: 55
Query Match: 9,48% Indels: 32
DB: 1 Gaps: 11

US-10-088-045-2 (1-502) x AF339049 (1-537)
Qy      177 VALTYRHISGVALAValAlaProValGlyMet-----HisIleAlaAsnGlyMet 193
Db      2 GCTATCATTTGC-----GCCCGAGGAAATGCTCCCTGGCAGCTGTTCCGGGAATG 55
Qy      194 TYRglYleuIleleuValgluProLysglulglYleuProLys----- 207
Db      56 CATGGAGCATATGTTGCTTCGCCAGAAATGGCTCAAGAAATCCGAGGGAAACCGCTG 115
Qy      208 ---ValAspLysGluTyrTyrValMetGingLysPheTyr-----ThrLys 222
Db      116 CACTACGACCGCATCTACTACATCGCGGAGAAATGATTTCTACATTCGAGAGCAGAAC 175
Qy      223 GLYLYSTYR-----GlygluIngLylleuInProPheAsp---Met 235
Db      176 GGAATATACAAACATTACATTCTCCGCGGAAAGCTTTGCCACACACGATATCATG 235
Qy      236 GluLysAlaIleArgglusPalaiglutryvalphasniglyserValglialaleu 255
Db      236 CGCAAGCTATC-----CGAATCATGTCGTTTAAAGCGCGCGTGGTGGCTC 286
Qy      256 ThrGlyLysAlaIleuValAlaLysValIglYgluThrValArgLeuPheValIglYasn 275
Db      287 ACCGCAAGAACGCAATGACGCGCAAGCTCGGTGAACCGTGT---ATGTTCCGTG---CAT 340
Qy      276 GlyglYProAsnLeuThrSerSerPheHisValIleGlygluIlePheAspLysValHis 295
Db      341 TCCGAAGCCCAACCGGACACCGCGCCGATATCATCGCGGCTTGGCACTACGTC--- 397
Qy      296 PhegluGlyglYlysglyglusn-----HisAsnIleGlnThrLeuIlePro 312
Db      398 TGGAGACCGGCAAGTTVAGCAACCCGCGGAAAGATCTGAGACCTCGTTATCCGG 457
Qy      313 AlaGlyValAlaAlaIleThrGluPheLysValAspValProGlyAspTyrValLeuVal 332
Db      458 GCGGATCAGCCGAGCGGCGCTTACACTTCGCGGACCGCGGCTTACGCTACGTC 517
Qy      333 AspHis 334
Db      518 AACCAAC 523

RESULT 36
AP003594/c 343250 bp DNA linear BCT 28-NOV-2001
LOCUS Nostoc sp. PCC 7120 DNA, complete genome, section 14/19.
DEFINITION AP003594 BA000019
ACCESSION AP003594.1 GI:17132824
VERSION
KEYWORDS
SOURCE
ORGANISM Nostoc sp. PCC 7120
            Nostoc sp. PCC 7120
            Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
REFERENCE 1
AUTHORS Kaneko, T., Nakamura, Y., Wolk, C.P., Kuritz, T., Sasamoto, S.,
            Watanabe, A., Iriuch, M., Ishikawa, A., Kawashima, K., Kimura, T.,
            Kishida, Y., Kohara, M., Matsumoto, M., Matsumoto, A., Muraki, A.,
            Nakazaki, N., Shimo, S., Sugimoto, M., Takezawa, M., Yamada, M.,
            Yasuda, M. and Tabata, S.
TITLE Complete genomic sequence of the filamentous nitrogen-fixing
JOURNAL Cyanobacterium Anabaena sp. strain PCC 7120
MEDLINE DNA Res. 8 (5), 205-213 (2001)
PUBMED 21595285
BASE COUNT 11759840
ORIGIN
REFERENCE 2 (bases 1 to 343250)
AUTHORS Kaneko, T.
TITLE Direct Submission

```

## JOURNAL

Submitted (02-MAY-2001) Takakazu Kaneko, Kazusa DNA Research Institute, The First Laboratory for Plant Gene Research, Yana 1532-3, Kisarazu, Chiba 292-0812, Japan  
(E-mail: kaneko@kazusa.or.jp,  
url: http://www.kazusa.or.jp/cyanobase/  
Tel: 81-438-52-3935 (ex. 2338), Fax: 81-438-52-3934)  
Location/Qualifiers

## FEATURES

## Source

1. .343250

/organism="Noctoc sp. PCC 7120"

/mol\_type="genomic DNA"

/db\_xref="taxon:103690"

/note="Synonym: Anabaena sp. PCC7120"

/complement (188. .6253)

/gene="a113691"

complement (188. .6253)

/gene="a113691"

/note="ORF\_ID: a113691"

/codon\_start=1

/transl\_table=1

/product="serine/threonine kinase with two-component sensor domain"

/protein\_id="BAB75390.1"

/db\_xref="GI:17132825"

/translation="MESVKNTYVDLAIAPGFIIEIYHSGKTVYRAVRADQDP  
VVIKLRKREYPTSELIQFRNOYAIANKIPGVINDYSLPYRNSYAMWEDFGIS  
LRDYAQOOSLITELITITITLADILHHIYQRIIHKDIKPAKILINPERKQVLDIF  
SIASLPERTONISPIHLEGTAYSEQOTGMRNGIDRSPYSGLVGFELTIGE  
LPTTDDPMELVCHIAKOPDKFTIPSCORAYANRNSGSESTKLPMNETIYQVGE  
IYKMLAKNAEDRYQNALGKTYDEKLAQLQETGKVEYTTISRDYCDFIIPDKY  
GREAEIETLLAARVSLGATEMLVAGFSGIKTAVINEKPIYARQRYFOGKED  
QFORNIPEFAVOAFRLMGOLLTESDTQIQWRSKYLEAVGRKGVILEVLELEI  
IGKOPAPKSLGTSLOIRPNLIFKKFPFTSAHPVILFDLOMADSGSLKLIQL  
MNDTGVLIGAVRDNVNPAPHMLTVNEIKGNNGITKIDLAFLNQLKNTLVAVET  
LKSEETASIALIVROKTOGNPFTTQPELKAHONLIDPNYDLGCMQCDIACVNHQ  
ALTDVYAFASQKLRPRATQEVQLAATGNOFDLETATYVESPMKTAQVMA  
LQEGLIIPQSEVKKFVGOEIEIDHAQTEITVYKFLHDRVQQAAYCLIPRARALAH  
HHIGQLLOKISPAAREHEIIEIYNOYNGITITITQOESRDELAQNLNAQARALAT  
AYQSEYADTSMTLLGETAMOROYQUTLTHELAIEIALCGDEQMOLITVVAH  
ACQPSRVOYQVOKIOLATSRNELLEIATGKSFLOELGVSLPDHPPEVQOAREE  
NRLIGRLIGEPINLPKMDPEKLAIMQISSIIPACVYMGSLIYLITVLOKLSIO  
FGNSLSAHGYVYAPHLSTTQNMALAOLOQIAYOLASEPRAKINRAATFVILGCT  
FYNCHSLQDTLPITQEGCTAGLETGCTFPGNVQVFGNNAWGSGLSEFSAQCT  
YHOQLDFNLVTSAMHCFYVWESALILIGSABEISLRQDAVEKIVTEALANDLWR  
LFOFLYLRVLNLADPRAKQDAEKARQYLTGCMGSVAEPIFYFYDSTALAEHQ  
PAGRSOMQVRENOALIDKADHAPMNYLHKQVLAETHRLLEOKTEAMEYEMAI  
KGAENKYLDELANELAKFYLDMGKEKVAVYVMOEAYCYARWEPARKTNDLEKR  
YPOLLOPILOERQPKRFQETIAVPGSSSTLSSTGTSISITPLDPASTILKAAOYIS  
TSLEBDELITNLTEIIEHNSGAKSALLPQEDIWQIKAWTLNFPQNSSESTQYLE  
SOPLTECEDIPKNIYYVNTQQTVDINLQTDIPGILGVMLOHQPQSVFCTPMNO  
GHLVGLIYENRLTRGVFTSDRLSEVIRLSAQAVSEKARLYQESQTKAOQKSLK  
OOKLFFNVNQMSOLDNAIFCVTONIRIRLIDVBRVGIYOPHLQVNYEGEFVAD  
VSPAPSAVALKQDHQCFGNVANYKQGRICAITVQVSESLDCHROLAOPHVAS  
LYVPIHQBELKGLICIHOCDRPQWEPLEMOQAQVGAOMGIALKQDILITQQA  
TQLEHTLOHLOQTQLOLVQNEKRSALGNVAGVAHEMNNPLGFAVSLQDTPPTQAI  
MEHLKYQESLPHRQDKILHAAQIDIDYLEDLPKLEAMVACDRLAKNISLSTLRF  
SRADKYKIPFNHKGIDSTILLKRLKANDRPAIEVTEYGNLPHICEFGOALD  
VFNNILANALDALEDANGLSVAEITANNRRIIRTOVANNHVRISITDGMISSEL  
KOHLEHLEFTTKGEKGTGLAIAIQIVVERKGALEVNSQLGCTGEPIITPTAB  
SANKY"

gene  
6804. .7658  
/gene="a13692"  
6804. .7658  
/gene="a13692"  
/note="ORF\_ID: a13692"  
hypothetical protein"  
/codon\_start=1  
/transl\_table=1  
/protein\_id="BAB75391.1"  
/db\_xref="GI:17132826"  
/translation="MNPPOIGVALLLTFQFKVQVGAITMVAQKILERPAKYR  
GLRTONIEPTLISYLVNIVSTRIVLIVAILDFGLETSPALALAAVGAIGAAMG  
GLIANPAGAFLLIFRPFKVGDFITNAGTIGTVESEGLFTTDTITPDNVLITVANKKI  
PADNIONFSANPFRVLDLALQHLHDVDHQAIALILKARISQIDNVIEENPAPDVEILTF

## gene

NLAGPYLAVRPYCNDHWQVYFDTNKAIREFTGEAGYPIPEQRYGVSPENGTSVV  
ISSALS"

## CDS

8079. .8939

/gene="a13693"

8079. .8939

/gene="a13693"

/note="ORF\_ID: a13693"

hypothetical protein"

/codon\_start=1

/transl\_table=1

/protein\_id="BAB75392.1"

/db\_xref="GI:17132827"

/translation="MNVABISQVALLLTFQGLRVGAIALMIIAQLINFGKILRR  
ARSKHVEPTLINVIVNIVDLIKLIVIAIIGFGEIETTSFALLAAGIAGAAMS  
GLIANPAGAFLLIFRPFVAGDTIKAGVAGKEIGLFTTINTSDNKTITGNKI  
PADNIONFSANPFRVLDLALQHLHDVDHQAIALILKARISQIDNVIEENPAPDVEILF  
NLTGTVLAVRPIYCHHEHYQVYFDTNKAIREFTGEAGYPIPEQRYAFTQOSPDTGSAN  
VQRTSLGRG"

## gene

9664. .9831

## CDS

/gene="a13694"

## CDS

9664. .9831

/gene="a13694"

/note="ORF\_ID: a13694"

unknown protein"

/codon\_start=1

/transl\_table=1

/protein\_id="BAB75393.1"

/db\_xref="GI:17132828"

/translation="MAVDLDRVLILQSFPHLRSWERNGLMEVSRRIKEHTVSM  
VAPEILVPG"

## gene

complement (9841. .11007)

## CDS

/gene="a113695"

complement (9841. .11007)

/gene="a113695"

/note="ORF\_ID: a113695"

/codon\_start=1

/transl\_table=1

/product="N-acyl-D-glucosamine 2-epimerase"

/protein\_id="BAB75394.1"

/db\_xref="GI:17132829"

/translation="MKNLQALALQLYKRALNDVLPFWENHSLDTDEGTYFTCLDRQK  
VYDIDFVILQNRQWTFMSLQNLQEKRENNMLKISNGKFLAQRHGRDAEGWYFALT  
RGSEPLVQVYNTFSDCFAMAFAFSOYLAKEBEMAKOVANQVANNVLRKNDPGKXTK  
TYGTEPMKALVPMPLIANLTLMEWLLEOETLENVSAIVQVNSDFLDKESGLAYE  
NVAPDGSHTDCEGRILNPGHGIEMWFMINDIARRNDKCTINOADVILNTKPHMD  
KVYQGIYVMDAAGHPQOQLEMDOKLWHLSEIYALMGVYLTGSEVWQEWYKQHD  
YSKQHPADDEYEMFGYLNRRGEVLNLKGRKKGCFHVPRALYLCWQPEALS"

complement (11134. .12273)

/gene="a113696"

complement (11134. .12273)

/gene="a113696"

complement (11134. .12273)

/note="ORF\_ID: a113696"

unknown protein"

/codon\_start=1

/transl\_table=1

/protein\_id="BAB75395.1"

/db\_xref="GI:17132830"

/translation="MLPRAIALDNGCRLLVAENGTQVLYKIQDOPVQVCFHTG  
GIYGVGVQDQKFTGLTGAGDAQNLVYNNNGSNCTDRKFSPEGKQMLRLIG  
LSLVENADADPKTDGAEVFTKYEQFLIDSSKRGQOLIYKAYTLNGFKYPODSRHIS  
PDTTFVRRINDERFLFLDPYGNALIEYFNSKQGNIIIPAGMFGKNIROGSAISG  
TWBPQEPQTKWIMRDRNDGAEDEKEEYRSDSSVTGMWDSKGVWTTIGDRGIG  
RHYPLQGLDNGNPITYTNSIOQEKTPLEFTDLRRIKYFPESDPTWYLSGFTAKNPFA  
EDPQAPSGSIACFDWMSKNRILRRIRIVPKOTIRKREMITASTH"

complement (12278. .13114)

/gene="a113697"

complement (12278. .13114)

/gene="a113697"

/note="ORF\_ID: a113697"

unknown protein"

/codon\_start=1

/transl\_table=1

/protein\_id="BAB75396.1"

/db\_xref="GI:17132831"

/translation="MNPPOIGVALLLTFQFKVQVGAITMVAQKILERPAKYR  
GLRTONIEPTLISYLVNIVSTRIVLIVAILDFGLETSPALALAAVGAIGAAMG  
GLIANPAGAFLLIFRPFKVGDFITNAGTIGTVESEGLFTTDTITPDNVLITVANKKI  
PADNIONFSANPFRVLDLALQHLHDVDHQAIALILKARISQIDNVIEENPAPDVEILTF

/gene="a113697"

/note="ORF\_ID: a113697"

unknown protein"

/codon\_start=1

/transl\_table=1

/protein\_id="BAB75396.1"





FAYFRTTRDGLSKENAKQRIQYVGRNELTSKALAEVNMKLGMIPTGSKORVDEL  
 QRETTIVSVECCSSSTGSLNSELMMEI PVQELVPGMIFLSGDAVPAVRITVADL  
 LNESVLTGNENIEKFECSYHLERKRFIPLKRMKDYNDLENNVCFKSHSVGNAX  
 AIVSTGKNTYSGLHTCCSRTS"  
 10496..11362  
 /locus\_tag="BC1582"  
 10496..11362  
 /locus\_tag="BC1582"  
 /BC\_number="3.1.11..-"  
 CDS  
 gene  
 Alignment Scores:  
 Pred. No.: 6,81e-05 Length: 347456  
 Score: 245.00 Matches: 108  
 Percent Similarity: 35.26% Conservative: 57  
 Best Local Similarity: 23.08% Mismatches: 159  
 Query Match: 9.33% Indels: 144  
 DB: 1 Gaps: 19  
 US-10-088-045-2 (1-502) x AE017003 (1-347456)  
 QY 67 ThrHsAlaProGluValProProValAspArgAspHisProAlaLysValValVal 86  
 DB 268075 ACTGAACAACGCTACACACCATTTGAAAGTTGAAAAAGACCAAGTTACTTTAATAGCG 268134  
 QY 87 LysMetGluThrValGluLysValMetArgLeuAlaAspGlyValGluTyrGlnPheTyr 106  
 DB 268135 AAA-----GAGAAAAAGCAAAATTAAAGTACCGTGTATTGTTCCAGCTCGG 268182  
 QY 107 ThrPheGlyGlyGlnValProGlyGlnMetLeuArgValArgGluGlyAspThrIleGlu 126  
 DB 268183 ACATTTAATGCTCATCTCTGCTTCAGAAATTCGGGTGAAAAAGGTGAAAAAGGTGAAA 268242  
 QY 127 ValGlnPheSerAsnHisProAspSerLysMetProHisAsnValAspPheHisAlaAla 146  
 DB 268243 GTGACCTTAAAAAT-----GAATTATCTGCACCGCATCTATTATTGGCATGCAATAT 268296  
 QY 147 ThrGlyProGlyGly-----GlyAlaGluAlaSerPheThrAlaProGly 161  
 DB 268297 CCTGTCCTCAATTAACATGATGATGATCCAGCGCTGACACAAATGCAATGTAACAGGAA 268356  
 QY 162 HisThrSerThrPheSerPheLysAlaLeuGlnProGlyLysTyrValTyrHisCysAla 181  
 DB 268357 AAAAGTTTCACTTACCAATTTGAAGGAACTTACAGGAACTTACGATCTGTCAT 268416  
 QY 182 ValAlaProValGlyMetHisIleAlaAsnGlyMetTyrGlyLeuIleLeuValGluPro 201  
 DB 268417 CAAGATTCTGTAAAT---CAATTAGCTAGAGGTTTGTAAGACCGCTCATTTGAGAAAT 268473  
 QY 202 LysGluGlyLeuProLysValAspLysGlyTyrTyrValMetGlnLysAspPhe----- 219  
 DB 268474 ACAAG-----GAAAAATATGATTAAGATTACACATTTATGTGATGATGATGATGAT 268527  
 QY 220 -----TyrThrLysGlyLysTyrGlyGlu--- 227  
 DB 268528 GATTAAGAAGAAATGATTAAGACGTTAAAGAAATACAAAGGCAATAGCTAATAA 268587  
 QY 227 ----- 227  
 DB 268588 TCTAAAGATTAAGAAATGGGCAAAAGATGATTAAGACCGCATGATCTTGAT 268647  
 QY 228 -----GlnGlyLeuGlnProPheAsp 234  
 DB 268648 ATGAACATGGGCACTGATAAAAAAGACTGATTAATATAGAGAGATGAGCAATGGAAT 268707  
 QY 235 MetGluLysAlaIleArgGluAspAlaGluTyrValPhe-----AsnGlySerVal 252  
 DB 268708 ATGAAGATGGAAGGTCATGATATGATGATGATGATTTTTCACATCAATGAGAAAAAGC 268767  
 QY 253 GlyAlaLeuThrGlyLysAsnAlaLeuLysAlaValGlyGluThrValArgLeuPhe 272  
 DB 268768 GGTGATTTTGTAGCG-----CATTAAGAAGTAAATAGGAGATTAACGTTGCTTTAGA 268821  
 QY 273 ValGlyAsnGlyGlyProAsnLeuThrSerSerPheHisValIleGlyGlu----- 289

DB 268822 CTCGTCAATGCTGT---TATCTATCATATATATACATGTTCAATGCGCATGATATMAA 268878  
 QY 290 -----IlePheAspLysValHis 295  
 DB 268879 GTAATTCGACAGATGCTACACCAATTAACGATCCAAAGTTATTAAGATTAAGTA--- 268935  
 QY 296 PheGluGlyGlyLysGlyGluAsnHisAsnIleGlnThrThrLeuIleProAlaGlyGly 315  
 DB 268936 ATTTCAATCGCACCGGTGTAACGTTATGATAT--- 268968  
 QY 316 AlaAlaIleThrGluPheLysValAspValProGlyAspTyrValLeuValAspHisAla 335  
 DB 268969 -----GAATTTACTCTATATATCTCGGAAATGATGTTGTAAGACCAT--- 269013  
 QY 336 IlePheArgAlaPheAsnLysGlyAlaLeuGlyIleLeuLysValGluGlyGluAsn 355  
 DB 269014 -----TCGAAAAATTAAGGTGCAAAAGAAATG--- 269040  
 QY 356 HisGluIleTyrSerHisLysGlnThrAspAlaValTyrLeuProGluGlyAlaProGln 375  
 DB 269041 -----AAAGCTGTTATTGATGATGATGATGATGATGATGATGATGATGATGAT 269073  
 QY 376 AlaIleAsp---ThrGlnGluAlaProLysThrProAlaProAlaAsnLeuGlnGluGln 394  
 DB 269074 ATGAAGAATTAACACAGACGAAAAAGAAATTAACGAAAGTAAATTAATGAAATATGAT 269133  
 QY 395 Ile-----LysAlaGlyLysAlaIleThrTyrAspSerAsnCysAlaAlaCysHisGlnProAsp 413  
 DB 269134 ACTAAAAAATTAGTATGATGTTTCAAGTAAATTAACGATATATGCTCCACATATATATATGAC 269193  
 QY 414 GlyLysGlyValProAsnAlaPheProProLeuAlaAsnSerAspTyrLeuAsnAlaAsp 433  
 DB 269194 -----TTAATATAGCGAA 269205  
 QY 434 HisAlaArgAlaAlaSerIleValAlaAsnGlyLysSerGlyLysIleThrValAsnGly 453  
 DB 269206 -----ATGAATGAAATTAAGATGATGATGATGATGATGATGATGATGATGAT 269241  
 QY 454 AsnGlnTyrGluSerValMetPro 461  
 DB 269242 AAGGATATTTCCGATATTAACCCA 269265  
 RESULT 38  
 AY078254 537 bp DNA linear BCT 03-MAR-2002  
 LOCUS Mesorhizobium sp. 4FB11 putative dissimilatory nitrite reductase  
 DEFINITION (nirK) gene, partial cds.  
 ACCESSION AY078254  
 VERSION AY078254.1 GI:19072874  
 KEYWORDS  
 SOURCE Mesorhizobium sp. 4FB11  
 ORGANISM Mesorhizobium sp. 4FB11  
 Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 Phyllobacteriaceae; Mesorhizobium.  
 REFERENCE 1 (bases 1 to 537)  
 AUTHORS Song, B. and Ward, B.B.  
 TITLE Nitrite reductase genes in halobenzoate degrading denitrifying  
 bacteria  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 537)  
 AUTHORS Song, B. and Ward, B.B.  
 TITLE Direct Submision  
 JOURNAL Submitted (08-FEB-2002) Geosciences, Princeton University, Guyot  
 Hall, Princeton, NJ 08544, USA  
 FEATURES  
 source  
 1..537  
 /organism="Mesorhizobium sp. 4FB11"  
 /mol\_type="genomic DNA"  
 /strain="4FB11"  
 /db\_xref="taxon:126151"  
 <1..>537  
 /gene="nirK"

CDS  
 <1..537  
 /gene="nitrK"  
 /codon\_start=2  
 /transl\_table=11  
 /product="putative dissimilatory nitrite reductase"  
 /protein\_id="AAL82513.1"  
 /db\_xref="GI:19072875"  
 /translation="VYHCAPEGMT PMHVSQMGTITWLPBDGLKNEKGLVKSPTVY  
 YIGDNEFYIPDEEDGNFKFSDSGSEYETMEVWNGLIPTHVNGRVLGTGDNAMK  
 SKYGETVLVPHSOANRSDRPHLIGGHGDYVWBEKFPANPKOLETWFIRIGSGAGAL  
 YTFROGVAVVNHNLIEA"

BASE COUNT 106 a 176 c 166 g 89 t

ORIGIN

Alignment Scores:  
 Pred. No.: 3,17e-07 Length: 537  
 Score: 232.00 Matches: 69  
 Percent Similarity: 52.78% Conservative: 26  
 Best Local Similarity: 38.33% Mismatches: 57  
 Query Match: 8.83% Indels: 28  
 Gaps: 10

US-10-088-045-2 (1-502) x AY078254 (1-537)

QY 177 ValTyrHisCysAlaValAlaProValGlyMet-----HisIleAlaAsnGlyMet 193  
 |||||  
 2 GTCTATCATTTGC-----GCGCGGAGGAGCATGATCCCTGGACACGTCTCGGGCATG 55  
 |||||  
 QY 194 TyrGlyLeuIleLeuValGluProGlyGluGlyLeuPro----- 206  
 |||||  
 56 CACGGCGCATATGATGCTGCTGCGCGGAGCATGCTCAAGAACGAGAGGCGCACTGTC 115  
 |||||  
 QY 207 LysValAspLysGluTyrTyrValMetGlnGlyAspPheTyrThrLysGlyLysTyrGly 226  
 |||||  
 116 AAGTCCGACACCGCTCTACTATGCGGATACGAGATTCTACATCCG---CGGACGAG 172  
 |||||  
 QY 227 GluGlnGlyLeuGlnProPheAsp--MetGluLysAlaIleArgGluAspAlaGlu-- 244  
 |||||  
 173 GACGGCAACTTCAAGAGTTCGACTCCCTGGCGAGCATACGAGAGCATGAGAGTC 232  
 |||||  
 QY 245 -----TyrValValPheAsnGlySerValGlyAlaLeuThrGly 257  
 |||||  
 233 ATGCGCGGCTGATCCGACTCAGCTGCTTCAACGCGCGGCTGCGGCTGACCGGC 292  
 |||||  
 QY 258 GluAsnAlaLeuLysAlaLysValGlyLysThrValArgLeuPheValGlyAsnGlyGly 277  
 |||||  
 293 GACACCGCATGAGTCCAGAGTGGCGGACGAGT---CTGTTCTGTC---CATCCACG 346  
 |||||  
 QY 278 ProAsnLeuThrSerPheHisValIleGlyGluIlePheAspLysValHisPheGlu 297  
 |||||  
 347 GCCAACCGGACCTCCAGGCCCACTGATCGCGCGCATGGCGCATGCTC---TGGGA 403  
 |||||  
 QY 298 GlyGlyLysGlyGluAsn-----HisAsnIleGlnThrThrLeuIleProAlaGly 314  
 |||||  
 404 GAGGCAAGATTCCGCAACCGCGCGCAAGAGTCTGAGACCTGTTCACTCGTGGCGGT 463  
 |||||  
 QY 315 GlyAlaIleIleThrGluPheLysValAspValProGlyAspTyrValLeuValAspHis 334  
 |||||  
 464 TCCGCGGCGCGGCGCTCTACACTTCTCGACGCGGCGCTTACGCTTACGCTACCAT 523  
 |||||

RESULT 39  
 AY078247 537 bp DNA linear BCT 03-MAR-2002  
 LOCUS Ensifer sp. 2FB8 putative dissimilatory nitrite reductase (nitrK)  
 DEFINITION  
 gene, partial cds.  
 ACCESSION AY078247  
 VERSION AY078247.1 GI:19072860  
 KEYWORDS  
 SOURCE Ensifer sp. 2FB8  
 ORGANISM Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 Rhizobiaceae; Sinorhizobium/Ensifer group; Ensifer.  
 REFERENCE 1 (bases 1 to 537)

AUTHORS Song, B. and Ward, B.B.  
 TITLE Nitrite reductase genes in halobenzoate degrading denitrifying  
 bacteria  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 537)  
 AUTHORS Song, B. and Ward, B.B.  
 TITLE Direct Submission  
 JOURNAL Submitted (08-FEB-2002) Geosciences, Princeton University, Guyot  
 Hall, Princeton, NJ 08544, USA  
 FEATURES  
 source  
 1..537  
 /organism="Ensifer sp. 2FB8"  
 /mol\_type="genomic DNA"  
 /strain="2FB8"  
 /db\_xref="taxon:126147"  
 <1..537  
 /gene="nitrK"  
 <1..537  
 /gene="nitrK"  
 /codon\_start=2  
 /transl\_table=11  
 /product="putative dissimilatory nitrite reductase"  
 /protein\_id="AAL82506.1"  
 /db\_xref="GI:19072861"  
 /translation="VYHCAPEGMT PMHVSQMGTITWLPBDGLKNEKGLVKSPTVY  
 YIGDNEFYIPDEEDGNFKFSDSGSEYETMEVWNGLIPTHVNGRVLGTGDNAMK  
 SKYGETVLVPHSOANRSDRPHLIGGHGDYVWBEKFPANPKOLETWFIRIGSGAGAL  
 YTFROGVAVVNHNLIEA"

BASE COUNT 106 a 174 c 170 g 87 t

ORIGIN

Alignment Scores:  
 Pred. No.: 5.03e-07 Length: 537  
 Score: 229.00 Matches: 70  
 Percent Similarity: 50.84% Conservative: 21  
 Best Local Similarity: 39.11% Mismatches: 62  
 Query Match: 8.72% Indels: 26  
 Gaps: 8

US-10-088-045-2 (1-502) x AY078247 (1-537)

QY 177 ValTyrHisCysAlaValAlaProValGlyMet-----HisIleAlaAsnGlyMet 193  
 |||||  
 2 GTCTATCATTTGC-----GACCGGAGGAGCATGCTCCCTGGACACGTCACTCGGGCATG 55  
 |||||  
 QY 194 TyrGlyLeuIleLeuValGluProGlyGluGlyLeuPro----- 206  
 |||||  
 56 AACGGCGCATATGATGCTGCTGCGCGGAGCATGAGTCAAGAGAGGAGCGCGCTG 115  
 |||||  
 QY 207 LysValAspLysGluTyrTyrValMetGlnGlyAspPheTyrThrLys----- 222  
 |||||  
 116 ACCTAGACACAGATCTACTATGTCGGCGGACGAGATTCTACTGCGCAAGACGAGCGCC 175  
 |||||  
 QY 223 -----GlyLysTyrGlyGlnGlyLeuGlnProPheAspMetGluLysAlaIle 239  
 |||||  
 176 GGGAACTACAGAGAGTCAAAACCCCGGCAAGCTTATGAAGATCTGTCAAGCGGATG 235  
 |||||  
 QY 240 ArgGlu---AspAlaGluTyrValValPheAsnGlySerValGlyAlaLeuThrGlyGlu 258  
 |||||  
 236 CGACGCTGACCCCGACCACTGCTTCAACGAGTGGCGGCGGCTGACCGGCGAGC 295  
 |||||  
 QY 259 AsnAlaLeuLysAlaLysValGlyGluThrValArgLeuPheValGlyAsnGlyGlyPro 278  
 |||||  
 296 CARGCTTTACGCGCGCGCTGGCGAG-----CGTGTCTCTGCTCTCCATTCGACGCGC 349  
 |||||  
 QY 279 AsnLeuThrSerSerPheHisValIleGlyGluIlePheAspLysValHisPheGluGly 298  
 |||||  
 350 AACCGGATACGAGCGCGGACCTGATCGGCGGCGATGACTATGTC---TGGCGGAC 406  
 |||||  
 QY 299 GlyLysGlyGluAsnHis-----AsnIleGlnThrThrLeuIleProAlaGlyGly 315  
 |||||  
 407 GGCAGTCCGCAACCGCGGATCTGACACGAGAAACCTGCTATTCGCGGCGGAGAC 466  
 |||||



```

QY      316 AlaAlaIleThrgluPheIysValAspValProGlyAspTyrValIleuValAspHis 334
DB      467 GCGGGGCGCTGCTTCTACACCTTCCGCGCACCGGGGTGTGTACCGCTACGTCACACAT 523

RESULT 40
AY078248                               537 bp   DNA       linear   BCT 03-MAR-2002
LOCUS      Ensifer sp. 4FB6 putative dissimilatory nitrite reductase (nirK)
DEFINITION
ACCESSION  AY078248
VERSION     AY078248.1  GI:19072862
KEYWORDS
SOURCE      Ensifer sp. 4FB6
ORGANISM    Ensifer sp. 4FB6
            Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
            Rhizobiaceae; Sinorhizobium/Ensifer group; Ensifer.
REFERENCE   1 (bases 1 to 537)
AUTHORS     Song, B. and Ward, B.B.
TITLE       Nitrite reductase genes in halobenzoate degrading denitrifying
            bacteria
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 537)
AUTHORS     Song, B. and Ward, B.B.
TITLE       Direct Submission
JOURNAL     Submitted (08-FEB-2002) Geosciences, Princeton University, Guyot
            Hall, Princeton, NJ 08544, USA
            Location/Qualifiers
FEATURES
     source          1..537
                     /organism="Ensifer sp. 4FB6"
                     /mol_type="genomic DNA"
                     /strain="4FB6"
                     /db_xref="taxon:126148"
                     <1..>537
                     /gene="nirK"
                     <1..>537
                     /gene="nirK"
                     /codon_start=2
                     /transl_table=11
                     /product="putative dissimilatory nitrite reductase"
                     /protein_id="AA182507.1"
                     /db_xref="GI:19072863"
                     /translation="VYHCAPEGVPMHWTSGMNGAIVLPDGLKDEKQPLTYDKIY
                     AAVGEQDFVYKQDAGNYKKYETBEDVAKMRTLTPTHTVFNQVVALTGDHALT
                     AAVGERVLVHISQANDTRPHLIGHGDDYVATGKFRNPDLDETTLIPGTAGAAAF
                     YTFRQGVVAVVHNLIEA"

BASE COUNT      106 a      175 c      170 g      86 t
ORIGIN
Alignment Scores:
Pred. No.:      5.86e-07      Length:      537
Score:          228.00      Matches:      70
Percent Similarity: 49.17%      Conservative: 19
Best Local Similarity: 38.67%      Mismatches: 62
Query Match:      8.68%      Indels:      30
DB:              1      Gaps:      8

US-10-088-045-2 (1-502) x AY078248 (1-537)
QY      177 ValTyrHisCysAlaValAlaProValGlyMet-----HisIleAlaAsnGlyMet 193
DB      2   GTCTATCATTTGC-----GCGCCGAGAGGCATGTGCTCGACACGTCACCTCGGGCATG 55

QY      194 TyrGlyIleuIleuValGluProLysGluGlyIleuPro-----                206
DB      56 AACGGGCGCATCATGTGCTGCGCGCGAGCTGAAGAGCAGAAAGGCCACGCGCTG 115

QY      207 LysValAspLysGlyTyrTyrValIleGlnGlyAspPheTyrThrLys-----        222
DB      116 ACGTACGACAAAGATCTACTATGTCTGCGACAGAGACTTCTACGTGCGGAGAGAGAGGCC 175

QY      223 -----GlyLysTyrGlyGluGlnGlyLeuGlnProPheAspMetGlyLysAlaIle 239
DB      176 GGGAACTACAAAGATGACAAACCCCGCGGAAAGCCTATGAAGATGCTGTCAAGGCGATG 235

```

```

QY      240 ArgGlu---AspAlaGluTyrValAlaPheAsnGlySerValGlyAlaLeuThrglyGlu 258
DB      236 CGACGCTGACCCCGACCCGACATCGTCTTCAACGGTGTGTGCGCGCGCTGACCGCGCAC 295

QY      259 AsnAlaLeuLysAlaLysValGlyGluThrValArgLeuPheValGlyAsnGlyGlyPro 278
DB      296 CATGCTTTGACTGCGCGCGCTGTGGCGAG-----CGTGTCTGTCGTCTCATTCGACAGGCC 349

QY      279 AsnLeuThrSerSerPheHisValIleGlyGluIlePheAspLysValHisPheGluGly 298
DB      350 AACCGGATACGCGACCGCATCGATCGCGGCGCATGTGACTATGTC-----GGG 400

QY      299 GlyLysGlyGluAsnHisAsn-----IleGlnThrThrIleuProAla 313
DB      401 GCGACCGGAAATTCGCGCAACCGCGGATCTGCACACAGAAACCTGCTCATTCGCGGGC 460

QY      314 GlyGlyAlaAlaIleThrgluPheLysValAspValProGlyAspTyrValIleuValAsp 333
DB      461 GGAAACCGCGCGGCTGCTTCTACACCTTCCGCGCACCGGGGTGTGTACGCTACGTCAC 520

QY      334 His 334
DB      521 CAC 523

```

Search completed: September 5, 2003, 12:57:48  
 Job time : 7408 secs

**THIS PAGE BLANK (USPTO)**





Best Local Similarity 36.3%; Pred. No. 5.8e-24; Matches 134; Conservative 45; Mismatches 149; Indels 41; Gaps 12;

QY 19 MLCGCSNOADKKAQPKSSTVDAAKTANADNAASOEHOGLPVIDAIVTHAPVPPVDR 78  
DB 12 MLGAGALAG--AAVAPLHTQAAHAGAAAAGAAPDISLPRVKDL-----VKPPFVH 64  
QY 79 DH-----PAKVVMKETEYKVRML-ADGVEYQFWTFGGQVPGQMIRREGDTIEVQFS 130  
DB 65 AHDQVAKTGPRVVEFTMTTEKKLVIREGTEIHAFTFNGSVGPIVMVHENDYVELRLI 124  
QY 131 NHDSKMPHNVDFHAATPGGGAASFTAGHSTSPFKLOGLVYHCAVAPVCM--- 187  
DB 125 NPDNTLHNDHIDHAAATGALGAGALTQVNGEETTLRFKATKPGVFVYHC--APEGVWPV 182  
QY 188 HIANGVYGLILVEPKGLP-----KVDREYVMOGDFYTK-----GKYGEQGLQFP 233  
DB 183 HVTSSNGAAMVLPFGGLGDEKQPLTYDKIYVGEQDFVPPDEAGNTKKTETPEAYE 242  
QY 234 DMEKAIRE-DAEYVNGSVGALTGENALAKAVGETVRLFVNGGPNLTSSFVIGEIPD 292  
DB 243 DAKKARTLPTHTIIVNGAVGALTGDHALTPAAVGE--RVLVHVSQANRDRPHLIGSHG 300  
QY 293 KYHFEKGKGNH---NIQTLLIPAGGAATFEKVDVPGDVLVDHAIFFAFKNGALGILK 349  
DB 301 YV-WATGKFRNPDPDLOETWTLIPGTAGAAFYTEROPGVAYVNHILIEAFELGAAGHFK 359  
QY 350 VEGEENHEI 358  
DB 360 VTGEMNDL 368

RESULT 5  
A48936  
nitrite reductase (EC 1.7.99.3) - Pseudomonas sp.

C:Species: Pseudomonas sp.  
C:Date: 19-Dec-1993 #sequence\_revision 18-Nov-1994 #text\_change 08-Oct-1999  
C:Accession: A48936  
R:Ye, R.W.; Fries, M.R.; Bezborodnikov, S.G.; Averill, B.A.; Tiedje, J.M.  
Appl. Environ. Microbiol. 59, 250-254, 1993  
A:Title: Characterization of the structural gene encoding a copper-containing nitrite reductase from Pseudomonas sp. strain A48936  
A:Reference number: A48936; PMID:93175864; PMID:8439151  
A:Accession: A48936  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-379 <YE1>  
A:Cross-references: GB:M97294; NID:g151396; PIDN:AACT9132.1; PID:g151397  
A:Experimental source: isolate G-179  
A:Note: sequence extracted from NCBI backbone (NCBIN:125554, NCBI:P.125555)  
C:Keywords: copper; oxidoreductase

Query Match 17.3%; Score 453; DB 2; Length 379;

Best Local Similarity 37.3%; Pred. No. 2.2e-23; Matches 118; Conservative 45; Mismatches 119; Indels 34; Gaps 10;

QY 72 VPPVVDNRH-----PAKVVMKETEYKVRMLAD-GVEYQFWTFGGQVPGQMIRREGD 123  
DB 59 VKRPFVHAHTQKAEGRPKVVEFKMTIOEKKIIVDDKTEVHAATFDGVSVPGRPMIYHDD 118  
QY 124 TIEVQFSNHPDSKMPHNVDFHAATPGGGAASFTAGHSTSPFKALOGVYHCAVA 183  
DB 119 YVELTLVNDPDTMELQHNDIFHSATGALGGALTIVNPGDTAVLRFKATKAGVFVYHC--A 176  
QY 184 PYGM---HIANGVYGLILVEPKGLP-----VKREYVMOGDFYTK-----GKYG 226  
DB 177 PAGMVPWHTVSGNNGAAMVLPFGGLGDEKQPLTYDKIYVGEQDFVPPDEAGNTKKTETPEAYE 236  
QY 227 EQLQFPDMEKAIRE-DAEYVNGSVGALTGENALAKAVGETVRLFVNGGPNLTSSFV 285  
DB 237 SAGEAIPDVLAEAKTILPTHTIIVNGAVGALTGDNALQAKVGD--RVLVHVSQANRDRPH 294  
QY 286 VIGELIDKYNHFEKGKGNH---NIQTLLIPAGGAATFEKVDVPGDVLVDHAIFFAFK 342

DB 295 LIGHGDDY-WATGKFRNPDPDLOETWTLIPGAGAAAYTFOQPGIYAVVNHILIEAFEL 353  
QY 343 GALGILKVEGEENHEI 358  
DB 354 GAAHGFKVTGDMNDL 369

RESULT 6

JG0170  
nitrite reductase (EC 1.7.99.3) - Alcaligenes denitrificans subsp. xylosoxydans  
C:Species: Alcaligenes denitrificans subsp. xylosoxydans  
C:Date: 23-Jul-1999 #sequence\_revision 23-Jul-1999 #text\_change 11-May-2000  
C:Accession: JG0170  
R:Suzuki, E.; Horikoshi, N.; Kohzuma, T.  
Biochem. Biophys. Res. Commun. 255, 427-431, 1999  
A:Title: Cloning, sequencing, and transcriptional studies of the gene encoding copper-catalyzed nitrite reductase from Alcaligenes denitrificans subsp. xylosoxydans  
A:Reference number: JG0170; PMID:99160880; PMID:10045725  
A:Accession: JG0170  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-360 <SUZ>  
A:Cross-references: GB:AF051831  
C:Keywords: copper; oxidoreductase

Query Match 17.2%; Score 451; DB 2; Length 360;  
Best Local Similarity 35.8%; Pred. No. 2.8e-23;  
Matches 134; Conservative 45; Mismatches 135; Indels 60; Gaps 15;

QY 41 AAKTANADNAASOEHOGLPVIDAIVTHAPVPP--PVDRHPAKVVMKETEYKVRMLA 98  
DB 18 AAGTMAODA-----DKLPHTKVTLVAPQVHPHQATKSGPKVVEFTMTTEKKMVID 71  
QY 99 D-GVEYQFWTFGGQVPGQMIRREGDTIEVQFSNHPDSKMPHNVDFHAATPGGGAAS 157  
DB 72 DKGITLQANTFNGSNPFPVTVHEDVYDVLVNPATNMPHNVDFHATGALGAKLTN 131  
QY 158 TAPGHTSPFKALOGVYHCAVAPVCM---HIANGVYGLILVEPKGLP-----K 207  
DB 132 VNPGEQATLRFADRSQTVYHC--APEGMVPMHVSQNGSGLTLPFGGLGDEKQPLH 189  
QY 208 VKREYVMOGDFY---TKGKGEQGL---QFPDMEKAIRE-DAEYVNGSVGALTGEN 259  
DB 190 YDRAVYTFEFDYIPKPGDKYATLAESYGDVQVWRTLTPSHIVNGKVALTGAN 249  
QY 260 ALKAVGETVRLFVNGGPNLTSSFVIGEIRDKYHFEKGKGNH---NIQTLLIPAG-- 314  
DB 250 ALTKAVGETVRLI--HSQANRDRPHLIGHGDMV-WETGKFRNPDPDLOETWTLIP 306  
QY 315 GAATFEKVDVPGDVLVDHAIFFAFKNGALGILKVEGEENHEIYSHKOTDAVYLP 374  
DB 307 GAAVYTFK--QPGVAYVNHILIEAFELGAAGHIVGKMN----- 346  
QY 375 QALDPOEAPKTPAP 388  
DB 347 ---DLMKQIKAPAP 357

RESULT 7

S32112  
(EC 1.7.2.1) (EC 1.7.2.1) - Pseudomonas aureofaciens  
C:Species: Pseudomonas aureofaciens  
C:Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 03-Jun-2002  
C:Accession: S32112  
R:Zumft, W.G.  
submitted to the EMBL Data Library, March 1993  
A:Reference number: S32112  
A:Accession: S32112  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-363 <ZUM>  
A:Cross-references: EMBL:Z21945; NID:g287906; PIDN:CAA79939.1; PID:g287907  
C:Keywords: copper; oxidoreductase



A:Residues: 1-455 <KUR>  
A:Cross-references: GB:AE007870, PIDN:AAK89058.1, PID:g15158854, GSPDB:GN00170  
C:Genetics:  
A:Gene: AGR\_L\_970  
A:Map position: linear chromosome

Query Match 16.8%; Score 442; DB 2; Length 455;  
Best Local Similarity 33.2%; Pred. No. 1,6e-22;  
Matches 122; Conservative 58; Mismatches 131; Indels 56; Gaps 12;

QY 21 SCCSNQADKAAOKPSTVDAAKATNAD-----NAAOHOGLPVIDAIVTHAEPV 73  
DB 106 ACAAEKKAAPKAPLTSALPAPRAKVDLVKPPVHAHTOKAEGGPKVEFTLT----- 159  
QY 74 PPVDRHAPKAVVKKMETVEKVRMLAD-GVEYQFWTFGGQVPGQMIRVREGDTIEVOFSNH 132  
DB 160 -----IK-----EQKMLDDKGTVEHAMTFNSVPGPLMVVHDDVVELTLINP 203  
QY 133 PDSKAPHVDFPAATGPGGGAASFTAPCHTSTFSFKALQPLVYHCAVAPVGM--HI 189  
DB 204 DTNELOHNIIDFHSATGALGGGGLITVNPGEKAILRFKATKAGVFVYHC--APPGVMPWHV 261  
QY 190 ANGWYGLILVEKEGRLP-----VDKEYVMQGFYTK-----GKYGEOGLQPFDM 235  
DB 262 TSGMNGALVLPREGLTGDKGKELYDYKVVYVGEODFYIPRDENGPFKYESAGAMADT 321  
QY 236 EKALIRE-DAEYVFNQSVGALTGENALKAKVGETVRLFPVNGGPNLTSSFFHIGEIFDKY 294  
DB 332 LEVWRKLTPIPSHIVFGAVGALTGHEALQAAGE--KVLIVHSQANRDRPHLIGHGIDVY 379  
QY 295 HPEGKGENH--NIQTLLIPAGGAALTEFKVDVPGDVLVDHAIFFRAFNKALGILKYE 351  
DB 380 -WATGKFRNPRLDDETFIFPGTGAAPFYTFEGGIYAVVNHNLIEAFELGAAAHFKYT 438  
QY 352 GEENHEI 358  
DB 439 GEMNNTL 445

RESULT 11  
A95347  
Probable nitrite reductase (EC 1.7.99.3) Nirx [imported] - Sinorhizobium meliloti (strain C) [Species: Sinorhizobium meliloti]  
C:Date: 24-Aug-2001 #sequence\_revision 24-Aug-2001 #text\_change 14-Sep-2001  
C:Accession: A95347  
R:Barrett, M.J.; Fisher, R.F.; Jones, T.; Komp, C.; Abola, A.P.; Barloy-Hubler, F.; Bove, P.; Kallman, S.; Keating, D.H.; Palm, C.; Peck, M.C.; Surzycki, R.; Wells, D.H.; Yen, K.C. Proc. Natl. Acad. Sci. U.S.A. 98, 9883-9888, 2001  
A:Title: Nucleotide sequence and predicted functions of the entire Sinorhizobium meliloti A:Reference number: A95347; MUID:21396509; PMID:11481432  
A:Accession: A95347  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-376 <KUR>  
A:Cross-references: GB:AE006469, PIDN:AAK65339.1, PID:g14523797, GSPDB:GN00165  
A:Experimental source: strain 1021, megaplasmid pSymA  
R:Galibert, F.; Pihan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler, P.; Chaint, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.; L.; Hyman, R.W.; Jones, T.  
Science 293, 668-672, 2001  
A:Authors: Kahn, D.; Kahn, M.L.; Kallman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure, M.; Vandenbol, M.; Vorholter, F.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.  
A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.  
A:Reference number: A96039; MUID:21368234; PMID:11474104  
A:Contents: annotation  
C:Genetics:  
A:Gene: nirx  
A:Genome: plasmid  
C:Keywords: oxidoreductase

Query Match 16.7%; Score 439.5; DB 2; Length 376;  
Best Local Similarity 33.2%; Pred. No. 1.8e-22;  
Matches 121; Conservative 52; Mismatches 128; Indels 63; Gaps 13;

QY 60 PVIDAIVTHAP-----VPPVDRDHP-----AKVVVKMETVE- 92  
DB 24 PLIGAVSAHAEAVAKTAHINVASLPRVKVDLVKPPVHAHTOKAEGGPKVEFTLTIEE 83  
QY 93 -KVRMLAGVEYQFWTFGGQVPGQMIRVREGDTIEVOFSNHDSKMPHNVDFHAATGPGG 151  
DB 84 KKIIVDEQGTIELHANTFNSVPGPLMVVHDDVVELTLINPTNTLLOHNIIDFHSATGALG 143  
QY 152 GAASFTAPCHTSTFSFKALQPLVYHCAVAPVGM--HIANGWYGLILVEPKGLP-- 206  
DB 144 GALTVPNPGDTTLRFPAASKAGVFVYHC--APPGVMPWHVTSGNMGALVLPREGLTIDG 201  
QY 207 -----KVDKEYVMQGFY-----TKKY-----GEOGLQPFDMKALIREDAEYVFNFS 251  
DB 202 KGNSTITVKKVYVGEODVYVRDANGKPKKYESGEAADTLEWRTL--TSHIVFNCA 259  
QY 252 VGALTGENALKAKVGETVRLFPVNGGPNLTSSFFHIGEIFDKVHPEGKGENH--NIQT 308  
DB 260 VGALTGDSALKAAAGE--KVLIVHSQANRDRPHLIGHGIDVY-WATGKFRNAPVDDET 316  
QY 309 TLIPAGGAALTEFKVDVPGDVLVDHAIFFRAFNKALGILKEGENHEIYSHKQTDAY 368  
DB 317 WFIPEGTGAAPFYTFEGGIYAVVNHNLIEAFELGAAAHFAVTGMDMDLMT-----SVR 371  
QY 369 LPEG 372  
DB 372 APSG 375

RESULT 12  
AC3633  
Nitrite reductase (EC 1.7.99.3) [imported] - Brucella melitensis (strain 16M)  
C:Species: Brucella melitensis  
C:Date: 01-Feb-2002 #sequence\_revision 01-Feb-2002 #text\_change 03-Jun-2002  
C:Accession: AC3633  
R:DeJvecchio, V.G.; Kaprat, V.; Redkar, R.J.; Patra, G.; Mujir, C.; Los, T.; Ivanova, .; Mazur, M.; Goltzman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Leless Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002  
A:Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis A:Reference number: AJ3252; PMID:11756688  
A:Accession: AC3633  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-376 <KUR>  
A:Cross-references: GB:AE008918, PIDN:AAL54230.1, PID:g17985202, GSPDB:GN00191  
A:Experimental source: strain 16M  
C:Genetics:  
A:Gene: BMEI10988  
A:Map position: II  
C:Keywords: oxidoreductase

Query Match 16.4%; Score 431; DB 2; Length 376;  
Best Local Similarity 34.8%; Pred. No. 6.9e-22;  
Matches 122; Conservative 48; Mismatches 147; Indels 34; Gaps 11;

QY 43 KTAANDNAAOHOGLPVIDAIVTHAEPVPPVDRHAPKAVVKKMETVEKVRMLAD-GV 101  
DB 38 RKASLEETIALPRLQ-KVELVDPPLVHA--HTQVAEGGPKVQFTVIEKKIIVDDAGT 93  
QY 102 EYQFWTFGGQVPGQMIRVREGDTIEVOFSNHDSKMPHNVDFHAATGPGGGAASFTAPG 161  
DB 94 EVHAMTFNGITVPGPLMVVHDDVVELTLINPTNTLHIDFHAATGALGGGGLTEINPG 153  
QY 162 HTSTFSFKALQPLVYHCAVAPVGM--HIANGWYGLILVEPKGLP-----KVDKE 211  
DB 154 EKTVLRFKATKPGVFVYHC--APPGVMPWHVTSGNMGAVMWLPREGLTGDKGNKLTVDKV 211  
QY 212 YVVMQGFYTK-----GKYGEOGLQPFDMKALIRE-DAEYVFNQSVGALTGENALKA 263  
DB 212 YVVGSDPYVRDENGKVTYEPADGDAVEDVTKWRTLTPHHVFNQSVGALTGENALKA 271  
QY 264 KVGETVRLFPVNGGPNLTSSFFHIGEIFDKVHPEGKGENHNI--QTTLLIPAGGAALTEF 321

Db 272 KXGKYLII--HSQANRDRPHLIGHGQYVWATGKFNTPPVDDGTWTFPGGAHAARY 329  
 322 KVDVPGDYVLVDHAFRAFNKAGLILKYEGEENHEIYHKOTDAVYLPFG 372  
 Db 330 TERPGIYAVVNHNLIEAFELGAAAHFYKTEGWNDLMT-----SILAPSG 375

## RESULT 13

139582

nitrite reductase (EC 1.7.99.3) [validated] - Alcaligenes faecalis

C:Species: Alcaligenes faecalis

C:Date: 19-Jul-1996 #sequence\_revision 19-Jul-1996 #ext\_change 15-Sep-2000

C:Accession: J39582

R:Author: M.; Suzuki, J.; Kukimoto, M.; Ohnuki, T.; Horiuchi, S.; Bepu, T.

J. Gen. Microbiol. 139, 725-733, 1993

A:Title: Cloning and characterization of nitrite reductase gene from Alcaligenes faecalis

A:Reference number: J39582; MUID:93294530; PMID:8515232

A:Accession: J39582

A:Status: preliminary; translated from GB/EMBL/DBD

A:Molecule type: DNA

A:Residues: 1-376 &lt;RES&gt;

A:Cross-references: GB:D13155; NID:G398121; PIDN:BA02440.1; PID:G398122

C:Genetics:

A:Gene: nlr

C:Keywords: oxidoreductase

F:311,172,181,186/Binding site: copper (His, Cys, His, Met) (type 1) #status experimental

F:316,171/Binding site: copper (His) (type 2) (shared with trimeric partner 2) #status experimental

F:342/Binding site: copper (His) (type 2) (shared with trimeric partner 1) #status experimental

Query Match 16.3%; Score 427.5; DB 2; Length 376;

Best Local Similarity 33.7%; Pred. No. 1.2e-21;

Matches 129; Conservative 51; Mismatches 158; Indels 45; Gaps 14;

Db 15 LSALMLSGCSNQADKAAQPKSSTVDA---AAKTANADNAASQEHGCELPVIDAIVTHAP 70

Db 13 LAGALAG-----ALAPVLTTSAMGQAVRKATTAETALPRQ-KVELVDPFVPHAH 64

Db 71 EVPPVDRDHPAAVYVMEVEKVMRLAD-GVEYQFTFGGQVPGCMIRREDTIEVQR 129

Db 65 S---QVAEGGPKVVEFTVVEKKIYVDAGTEVHMAAFNGTVPGPLMVVHODDYELTL 121

Db 130 SNHPSMPHNVDFHAATGPGGGAASFTAPGHTSTFSFKALQGLVYVHCAVAPVGM--187

Db 122 INEPTTILMNIIPHAATGALGGGGLTEINPGEKTIIRFQATRGVYVHC--APGGMV 179

Db 188 -HIANGVYGLIVEPEKLP-----KVDKEYVVMGDFYT---KGKYE--OGLQR 232

Db 180 MHWVSGNGAIVMLPRGLHDGKGAALTYKIVYVGEQDFVPRDEGKTKYEAPQDAY 239

Db 233 PDMEKAIRE-DAEYVYVNGSVGALTGENALAKAVGETVRLFVNGGPNLTSSFHVIGEIR 291

Db 240 EDTVAKVIRTLPTTHVENVGAVALTGDKAMTAANG--KYLIVHSAQNRDTRPRLIGGCG 297

Db 292 DKVHPEGKGENNI--QTLIPAGSAITEFKVDVGDVVLVDHAFRAFNKAGLILK 349

Db 298 DYVWATGKFNTPPVDDGTWTFPGGAHAARYTQPGIYAVVNHNLIEAFELGAAAHFK 357

Db 350 VEGEENHEIYSHKOTDAVYLPFG 372

Db 358 VTGEMNDLMT-----SVLAPSG 375

## RESULT 14

E84274

membrane protein [imported] - Halobacterium sp. NRC-1

C:Species: Halobacterium sp. NRC-1

C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #ext\_change 02-Feb-2001

C:Accession: E84274

R:Author: W.V.; Kennedy, S.P.; Mahirae, G.G.; Bergquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.

J. Leitauer, B.; Keller, K.; Cruz, R.; Danon, M.; Hough, D.W.; Maddocks, D.G.; Jabcic

Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000

A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Li

A:Title: Genome sequence of Halobacterium species NRC-1.

A:Reference number: A84160; MUID:20504483; PMID:11016950

A:Accession: E84274

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-379 &lt;STO&gt;

A:Cross-references: GB:AE004437; NID:G10580721; PIDN:AA019561.1; GSPDB:GN00138

C:Genetics:

A:Gene: pan1

Query Match 9.5%; Score 250.5; DB 2; Length 379;

Best Local Similarity 25.3%; Pred. No. 1.4e-09;

Matches 94; Conservative 52; Mismatches 133; Indels 93; Gaps 20;

Db 20 LSGCSNQADKAAQPKSSTVDAAR---TANADNAASQEHGCELPVIDAIVTHAPEVPP 75

Db 29 IAGCT--APDGESEYTDITTAQKOSGLPTSPPEVDAITEGQNVTLKSV-----PA 78

Db 76 VDRDHPAAVYV--VMEVEKVMRLADGVEYQFTFGGO-----VPGCMIRVEGDTIEV 127

Db 79 VADVHPILDSMGSPVQLPRV-----WAFATEDGDRSPGPIVREBQDILEV 124

Db 128 QFSNHPDSKMPHNVDFHAATGPGGGAASFT-----TADGHTSTFSFKALQGL 175

Db 125 TLDN-TDGRPHTLHFH-----GSQTAMEDDGVPTTGRVSGEKHTYIIPANVGT 176

Db 176 VYHCAVAPVGMHIANGVYGLIVEPEKGLPRVDKEYVVMGDFYTKKYGSGQIOPDM 235

Db 177 HLHYCHY-OTQRHIDMGVGFIRIDPK-GYEPADKEYFTVVDWDSLRSMAG----- 228

Db 236 EKAIREDAEVY--VNGSVGALTGENA-----LKAKEVTEVLFVNGGPNLTS 282

Db 229 -----EDVDYSRTNPNPVFTYNGKSAPRTLHPEDGSPITVBOGTVLHLVNG-YNMH 282

Db 283 SFHVIGEIFDKVHEGCK--GENNIO-TLLIPAGSAITEFKVDV-PGDIYL---VD 333

Db 283 PLHINHRFRVYKDGQIPAAQHTMTDTPMAPERRHTI-EFOADAPGIVLHCHKVN 341

Db 334 HAFRAFNKAGL 345

Db 342 HWNGTFYPGGM 353

## RESULT 15

AG2298

hypothetical protein all3942 [imported] - Nostoc sp. (strain PCC 7120)

C:Species: Nostoc sp. PCC 7120

A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120

C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #ext\_change 09-Dec-2002

C:Accession: AG2298

R:Author: T.; Nakamura, Y.; Molk, C.P.; Kuritz, T.; Sasamoto, S.; Matanabe, A.; Iriyuchi

Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S

DNA Res. 8, 205-213, 2001

A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana

A:Reference number: AB1807; MUID:21595285; PMID:11759840

A:Accession: AG2298

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-338 &lt;KUR&gt;

A:Cross-references: GB:BA000019; PIDN:BA075641.1; PID:G17133076; GSPDB:GN00179

C:Experimental source: strain PCC 7120

C:Genetics:

A:Gene: all3942

Query Match 9.5%; Score 248.5; DB 2; Length 338;

Best Local Similarity 27.8%; Pred. No. 1.6e-09;

Matches 76; Conservative 42; Mismatches 116; Indels 39; Gaps 12;

Db 94 VKRLAGVEYQFTFGGQVPGCMIRREDTIEVQSNHPDSKMPHNVDFHAATGSGGA 153

Db 93 VIOLNSAVSYNIWDLNIGRIIPGTLRAKQDRIRVLV--HNQAGSHSLHFHGV---HPA 146



QY 154 EASFTAP---GHSTSFKALQGLVYHCAVAPVGMHIANGVGLIVEPKKGLPKVDK 210  
C/Genetics: NMA0919; NMA0925  
A:Gene: NMA0919; NMA0925  
C:Superfamily: T. aquaticus cytochrome c552; cytochrome c6 homology  
C:Keywords: chromoprotein; electron transfer; heme; iron; metalloprotein  
F:1-21/Domain: signal sequence #status predicted <SIG>  
F:22-163/Product: cytochrome c552 #status predicted <MAT>  
F:28-109/Domain: cytochrome c6 homology <CY6>  
F:38-41/Binding site: heme (Cys) (covalent) #status predicted  
F:42-92/Binding site: heme iron (His, Met) (axial ligands) #status predicted

Db 147 EMDGVFVSNNSATIEFPAEPVHLYHCHIEPVTRHAKGLYGMFIIDPPRPBPABE 206  
QY 211 EYVMGQDFYTKGKGEGQLQPFDMKALREDAEYVYVNGVALGEMALKKAVGETVR 270  
Db 207 IYLVNMG-----YVDNNSHND--FAFNLPHHYM-DNEIQYQNLK 248  
QY 271 LFGVN---GGPNLTSEFHVIGELFDKVFHEGKGENHNIQTLLIPAGAA--ITFEKVDV 325  
Db 249 LYLNIIEEDPAT--FHLHAFEDVYR--GMSMKASEKTDVITMGVNERHLEAFR 304  
QY 326 PGDYLVNDAHFAFPAKGLIKLV--EGEENH 356  
Db 305 PGKMFPHODAIEN-GCMQPEVAVANNQNH 336

RESULT 16  
cytochrome c552 NMB0717 precursor [similarity] - Neisseria meningitidis (strain MC58 ser C:Species: Neisseria meningitidis  
C:Date: 01-Sep-2000 #sequence\_revision 01-Sep-2000 #text\_change 19-Jan-2001  
C:Accession: B81167  
R:Retcelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A. Hickey, E.K.; Hart, D.H.; Salberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.; Xi, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignan, V.; Pizzi, M.  
Science 287, 1809-1815, 2000  
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappoli, R.; V. A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.  
A:Reference number: A81000; MUID:2015755; PMID:10710307  
A:Accession: B81167  
A:Molecule type: DNA  
A:Residues: 1-138 <TET>  
A:Cross-references: GB:AE002426; GB:AE002098; NID:G7225939; PIDN:ANF41130.1; PID:G722594  
A:Experimental source: serogroup B, strain MC58  
C:Genetics:  
A:Gene: NMB0717  
C:Superfamily: T. aquaticus cytochrome c552; cytochrome c6 homology  
C:Keywords: chromoprotein; electron transfer; heme; iron; metalloprotein  
F:1-26/Domain: signal sequence #status predicted <SIG>  
F:27-138/Product: cytochrome c552 #status predicted <MAT>  
F:33-114/Domain: cytochrome c6 homology <CY6>  
F:43-46/Binding site: heme (Cys) (covalent) #status predicted  
F:47-97/Binding site: heme iron (His, Met) (axial ligands) #status predicted

Query Match 8.5%; Score 222; DB 1; Length 138;  
Best Local Similarity 42.7%; Pred. No. 3.1e-08;  
Matches 44; Conservative 14; Mismatches 45; Indels 0; Gaps 0;

QY 398 GRATYDSNCAACHQPDGKGVNAPFPLANSDYLNADHAAASIVANGLSGKITVNGNOYE 457  
Db 35 GQKVESNCAVCHGKGGEGRGTMFPPLYSDFIMKKPQVLLHSMVGVINGINTIKVNGKTYN 94  
QY 458 SYMPAIALSDQOIANVITTLNSFGKGGQSLADDAKAKTK 500  
Db 95 GMPATAIISDADIAAVATYIMAFDNGSGSVTEKDYKQAKSK 137

RESULT 17  
cytochrome c552 NMA0925 precursor [similarity] - Neisseria meningitidis (strain Z2491 se C:Species: Neisseria meningitidis  
C:Date: 01-Sep-2000 #sequence\_revision 01-Sep-2000 #text\_change 02-Feb-2001  
C:Accession: F81938  
R:Parhill, J.; Achman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel, J.; Holroyd, S.; Jagers, K.; Leach, S.; Mout, S.; Mungall, K.; Quail, M.A.; Rajandream, Nature 404, 502-506, 2000.  
A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.  
A:Reference number: A81775; MUID:20222556; PMID:10761919  
A:Accession: F81938  
A:Molecule type: DNA  
A:Residues: 1-163 <PAR>  
A:Cross-references: GB:AL162754; GB:AL157959; NID:G7379424; PIDN:CAB84197.1; PID:G737963

A:Experimental source: serogroup A, strain Z2491  
C:Genetics:  
A:Gene: NMA0919; NMA0925  
C:Superfamily: T. aquaticus cytochrome c552; cytochrome c6 homology  
C:Keywords: chromoprotein; electron transfer; heme; iron; metalloprotein  
F:1-21/Domain: signal sequence #status predicted <SIG>  
F:22-163/Product: cytochrome c552 #status predicted <MAT>  
F:28-109/Domain: cytochrome c6 homology <CY6>  
F:38-41/Binding site: heme (Cys) (covalent) #status predicted  
F:42-92/Binding site: heme iron (His, Met) (axial ligands) #status predicted

Query Match 8.5%; Score 222; DB 1; Length 163;  
Best Local Similarity 42.7%; Pred. No. 3.9e-08;  
Matches 44; Conservative 14; Mismatches 45; Indels 0; Gaps 0;

QY 398 GRATYDSNCAACHQPDGKGVNAPFPLANSDYLNADHAAASIVANGLSGKITVNGNOYE 457  
Db 30 GQKVESNCAVCHGKGGEGRGTMFPPLYSDFIMKKPQVLLHSMVGVINGINTIKVNGKTYN 89  
QY 458 SYMPAIALSDQOIANVITTLNSFGKGGQSLADDAKAKTK 500  
Db 90 GMPATAIISDADIAAVATYIMAFDNGSGSVTEKDYKQAKSK 132

RESULT 18  
copper resistance protein pcoa precursor - Escherichia coli plasmid pRJ1004 S52253  
N:Alternate names: copper-binding protein  
C:Species: Escherichia coli  
C:Date: 07-May-1995 #sequence\_revision 21-Jul-1995 #text\_change 11-Jun-1999  
C:Accession: S70159; S52253  
R:Brown, N.L.; Barrett, S.R.; Camakaris, J.; Lee, B.T.O.; Rouch, D.A. Mo. Microbiol. 17, 1153-1166, 1995  
A:Title: Molecular genetics and transport analysis of the copper-resistance determinant A:Reference number: S70159; MUID:96130847; PMID:8594334  
A:Accession: S70159  
A:Status: nucleic acid sequence not shown  
A:Molecule type: DNA  
A:Residues: 1-605 <BRO>  
A:Cross-references: EMBL:X83541; NID:G619126; PIDN:CA58525.1; PID:G619128  
C:Genetics:  
A:Gene: pcoa  
A:Genome: plasmid pRJ1004  
C:Superfamily: laccase  
C:Keywords: copper binding  
F:1-33/Domain: signal sequence #status predicted <SIG>  
F:33-605/Product: copper resistance protein pcoa #status predicted <MAT>

Query Match 7.7%; Score 203.5; DB 2; Length 605;  
Best Local Similarity 22.4%; Pred. No. 4e-06;  
Matches 89; Conservative 58; Mismatches 167; Indels 83; Gaps 14;

QY 100 GYEVQFWTFGGVPCQMIRVREGDTIEVQFSNHPDSK-----MPHNVFHAATGCG 150  
Db 60 GSERQAKTINGLPGPVLMKXEGDTITLVKVRKLEQTSIHNGIILPANMD----- 111  
QY 151 GGAASFTA--RGHSTSFKALQGLVYHCAVAPVGMHIANGVGLIVEPKKGLP-K 207  
Db 112 GVPGLSEFGIEPDDVYVTFPVKQNGTYWH--SHSGLOEOEGVYGAIIIDAREBPBA 168  
QY 208 VDEYVVMGQDFYTKGKGEGQLQPFDMKALREDAEYVYVNGVALGEMALKKAVGETVR 263  
Db 169 YREHVVMLSDVTE-----NHSHLLKLLKQSDVYNFRKPVGSRFRDVRNRLGSLA 220  
QY 264 KYGETVRLFVNGGPNLTSEFHVIGELFDKVFHEGKGENHNIQTLLIPA-----CG 315  
Db 221 TTAAD--RKWAMEKKNPTDLADVSG--YTYUULNMGQAPLKMWTGLFRPGEKIRLRFING 276  
QY 316 AATFEKVDVPGDYLVNDAHFAFPAKGLIKLV--EGEENHIIYSHKQTDAYL----- 369  
Db 277 SAMTFEDIRIPGLKMTVVAADGQYVNPVTFDEFRIAVAEYDVIYVPEGEATTTIQAQMD 336  
QY 370 -----PEGAPOAIDTOEAPKTPAPANLQEQIKAKATYD-----SNC 406

Db 337 RTGVARTLATRGLSLAAV-----PRLDPRLLTMEQMGCMGHDAAGMDHSGMGMDNS 392  
 QY 407 AACHPDGGKVPN---AFPLPANSDYLNADHAPAAST 440  
 Db 393 GEMMSMDGADLPDSCGTSAPMDHSHSMAGMDHSHRAGM 429

RESULT 19  
 B83910  
 hypothetical protein BH2082 [imported] - Bacillus halodurans (strain C-125)  
 C/Species: Bacillus halodurans  
 C/Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 15-Jun-2001  
 C/Accession: B83910  
 R/Takaki, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; Hirai  
 Nucleic Acids Res. 28, 4317-4331, 2000  
 A/Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and  
 A/Reference number: A83650; MUID:20512582; PMID:11058132  
 A/Accession: B83910  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-500 <STO>  
 A/Cross-references: GB:AP001514; GB:BA000004; NID:910174613; PIDN:BA05801.1; GSPDB:GN00  
 A/Experimental source: strain C-125  
 C/Genetics:  
 A/Genes: BH2082

Query Match 7.4%; Score 194; DB 2; Length 500;  
 Best Local Similarity 20.7%; Pred. No. 1.4e-05;  
 Matches 105; Conservative 71; Mismatches 198; Indels 134; Gaps 23;

QY 20 LSGCSNQAADKAAPKSTVDAAKTANADNAASOEHQELPVIDAIVTAPEVPPDRD 79  
 Db 18 IAGCS-----DDHSEMDHGNQOEBQNEGT----- 44

QY 80 HPAKVVMKETEVEKWLADGVEYQFTFGGOVPGQWIRREGDTLEVQSNHPSKMP 139  
 Db 45 EDKQKVFNLVEITHMFMFNEVMDAMTNGTLPQGEIRVOEGDEVIVANK--SLNVPT 102

QY 140 NVDPHAATPG---GGAESTAPAGHTSTFSPKALQPLYYHCVAVAVGMIHANGY 194  
 Db 103 ALHIGFPPVNMDCVGVTONAIMPGEFTYEYQANVPCTYYH--SHDQATQVVRGLY 161

QY 195 GLTIVREKGLPKVDKEYY---VMQDPEYTG---KQGE---QGLQPRME 236  
 Db 162 GVIVETP-----VDPESTDTDAIVIDEFASMGCMGMDHSHMDHGDMDSHS 215

QY 237 KAIREDAEYVVFNGSVGALTGENALKAKVGETVRLFVNGCGPMLTSSFHVIGEIFDKVF 296  
 Db 216 EMNEMWDTVINGKADPQI--ESINNVGGRVRLRVNAG--LFTQVVISPEHSFKITTH 272

QY 297 EGGKGEHNIQTLLIPAGAAITFEKVDVEDY---VLVDHAFRANKKALGILKKEG 353  
 Db 273 DGPVNEPVDVLOD-----TAFRIAPARYDLEITMD-----NPGAMGI-QVFAE 315

QY 354 ENHE-----IYSHKQDAVVLPEGAPOAIDQBAKTPAPANLQOKAKGAT--YD 403  
 Db 316 ENDRRLALIPLYVEEYDEEL-----QTVDSISSFDLTITGIRKELDIDGITKEYD 368

QY 404 SNCAACHQPD---GKGVNP-----AFPLANSDY-----LNADHARAASIV 441  
 Db 369 MLGSDGCGDTFTINGKKFPHMEIYDVEGDMVFTIINDPDPHMLHGDFFHVISKG 428

QY 442 ANGLSG---KITVN---GNOYESVMPA 462  
 Db 429 GKAIQGSPLVKDTLNVPRGETYEIVFEA 456

RESULT 20  
 A98239  
 hypothetical protein AGR\_L1722 [imported] - Agrobacterium tumefaciens (strain C58, Cere  
 C/Species: Agrobacterium tumefaciens  
 C/Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 18-Nov-2002

C/Accession: A98239  
 R/Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,  
 A.; Liu, F.; Mollam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappas, C.; Markelz, B.,  
 Science 294, 2323-2328, 2001  
 A/Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum  
 A/Reference number: A97359; MUID:21608551; PMID:11743194  
 A/Accession: A98239  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-449 <KUR>  
 A/Cross-references: GB:AE007870; PIDN:AAK89435.1; PID:915159296; GSPDB:GN00170  
 C/Genetics:  
 A/Genes: AGR\_L1722  
 A/Map position: linear chromosome

Query Match 7.1%; Score 187; DB 2; Length 449;  
 Best Local Similarity 21.2%; Pred. No. 3.6e-05;  
 Matches 97; Conservative 63; Mismatches 175; Indels 122; Gaps 20;

QY 12 ICALSLMLGCSNQAADKAAPKSTVDAAKTANADNAASOEHQELPVIDAIVTAPE 71  
 Db 7 LIGVSAAMVSSA-----AMATSSNS-----LPEALMETATQAP- 42

QY 72 VPPVDRDHPAKVVK-----METYEKVRLLADGVEYQFTFGGQVP 113  
 Db 43 VKPTGPDVNPVVTINGWTLPPRMNNGVKEFHLVAEPVEREM--AEGMTARLMGYGQSP 100

QY 114 GQMIRVRREGDTLEVQSNHPSKMPHN--VDHFAATPG---GGAESTAPAGHTSNF 166  
 Db 101 GPTTEAVEEDRVRIFVTN---KLPEHTTHHGKILPBGMDGVGLSOPHIIVGKTFVY 156

QY 167 SPKALQPLYYHCVAVAVGMIHANGYGLTIVREK--GLPKVDKEYVMQDPEYTKGY 225  
 Db 157 EBDLVKSGTFMTH--PHSDMVGAMGMKGFYIHPDPAFMVVDHDFVLNAF----- 209

QY 226 GEGQLQPRMEKAIREDAEYVVFNGSVGALTGENALKAKVGETVRLFVNGCGPMLTSSPH 285  
 Db 210 ---DINPGASVPKIMTMTDFNLAMNSRVFPGIDPLVSKNBRVVRVG---NLMTMTH 262

QY 286 VI---GEIDKXHPFG--KGENHNIQTLL--IPAGAAITFEKVDVDPGQVYLVDHAFRA 339  
 Db 263 PLHMHYDEVTCTDGVWRPAPRMEVSIDIIPVGAMRAYEEDPAKYGSMALHCHSHHT 322

QY 340 FNKGALGILKVEGENHEIYSHKQDAV-----YLPEGAPOAID---TOEAP 383  
 Db 323 KN--AMG-----HDVPTPIGVDKTVTEKIRKIRDPYMPGTAGMADGAMEMEIRP 371

QY 384 KTRPAP-----ANLQOKIAGKAT 401  
 Db 372 ENTVPMTGTGMPHPIEMGMEFSVVKREGISAGDYT 408

RESULT 21  
 AC3047  
 multicopper oxidase [imported] - Agrobacterium tumefaciens (strain C58, Dupont)  
 C/Species: Agrobacterium tumefaciens  
 C/Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 18-Nov-2002  
 C/Accession: AC3047  
 R/Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chan, L.; Wood, G.E.; Chen, Y.; Woo, I  
 erage, G.; Gille, W.; Grant, C.; Guenther, D.; Kutyavlin, T.; Levy, R.; Li, M.; McCell  
 i Karp, P.; Romero, P.; Zhang, S.  
 Science 294, 2317-2323, 2001  
 A/Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,  
 ster, E.W.  
 A/Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
 A/Reference number: AB2577; MUID:21608550; PMID:11743193  
 A/Accession: AC3047  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-449 <KUR>  
 A/Cross-references: GB:AE008669; PIDN:AAU44793.1; PID:917742433; GSPDB:GN00187  
 A/Experimental source: strain C58 (Dupont)  
 C/Genetics:





R.Nieman, W.C.; Feldblum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.  
B.; Laid, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kojan  
n, J.; Kholmoeva, M.; White, O.; Salzberg, S.L.; Shapito, L.; Venter, J.C.; Fraser, C.M.  
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A.Title: Complete Genome Sequence of *Calobacter crescentus*.

A.Reference number: A87249; MUID:21173698; PMID:11259647

A.Accession: H87368

A.Status: preliminary

A.Molecule type: DNA

A.Residues: 1-570 <STO>

A.Cross-References: GB:AE005673; NID:g13422242; PIDN:AAK2948.1; GSPDB:GN00148

C.Genetics:

A.Gene: CC0964

Query Match 6.2%; Score 163; DB 2; Length 570;

Best Local Similarity 25.2%; Pred. No. 0.0021;

Matches 69; Conservative 45; Mismatches 112; Indels 48; Gaps 14;

QY 73 PPVDRDHPAKVYVMEYKVRMLADGVEYQFWTGGQVPGQMIRVREGDTIEVQFSNH 132  
DB 45 PPALSGEE-----IKL-TVGHMAKIDGKAGHAVVNGAIPGLIRLKEGQVRLSVTNT 98

QY 133 PDSKPHNVDFHAAACP---GGGAASFTA--FGHTSTFSFKALQPLGVYVHCAPVGM 187  
DB 99 LDEDT--SIHMGGLVFPQMDVPGVSPGIRPGETFYEPFIRSGTYVMH--SHSGL 153

QY 188 HIANGMYGLIIVEPK-EGLPKVDKEYVYQGDFTYTKGKYGQGLQPFDEKAIREDAEVY 246  
DB 154 QEQMHYGMIIIDPAGEDVAVNDREHVVLSDM-----SFLPHLEFRLTKXSAGH- 204

QY 247 VFNGS---VGAITGEN-ALKAKVGETVRLFYNGGPNLTSSFFHVIIGEIFDKVHREGKG 301  
DB 205 -FNROKQTVAGILKQDQSLKRVRA-----WGQRMMDPTDIADTVGVSYYVLIANGHGP 257

QY 302 EN-----HNIQTLLPAGGAATTEKVDVPG 327  
DB 258 DNMGLFAPGERVRLRFINAG--AMTTFNVRIPG 289

# RESULT 27

S11027 L-ascorbate oxidase (EC 1.10.3.3) precursor - Cucurbita cv. Eblisu Nankin

C.Species: Cucurbita cv. Eblisu Nankin

C.Date: 21-Nov-1993 #sequence\_revision 10-Nov-1995 #text\_change 23-Sep-2002

C.Accession: S11027; S36936

R.Esaka, M.; Hattori, T.; Fujisawa, K.; Sakajo, S.; Asahi, T.

Eur. J. Biochem. 191, 537-541, 1990

A.Title: Molecular cloning and nucleotide sequence of full-length cDNA for ascorbate oxi

A.Reference number: S11027; MUID:9031033; PMID:2143984

A.Accession: S11027

A.Molecule type: mRNA

A.Residues: 1-579 <EUR>

A.Cross-References: EMBL:X55779; NID:g18251; PIDN:CAA39300.1; PID:g18252

A.Accession: S36936

A.Molecule type: protein

A.Residues: 31-48 <ESA>

C.Superfamily: laccase

C.Keywords: oxidoreductase

F.373-566/Domain: carboxyl-terminal beta-barrel #status predicted <BB3>

F.49-231/Disulfide bonds: #status predicted

F.90-478/Binding site: copper (His) (type 2) #status predicted

F.192-134,480,536,538/Binding site: 2Cu-O cluster (His) (copper type 3) #status predi

F.111-568/Disulfide bonds: #status predicted

F.193,397,473,542/Binding site: substrate (Trp, Trp, Glu, His) #status predicted

F.210-222/Disulfide bonds: #status predicted

Query Match 6.2%; Score 163; DB 2; Length 579;

Best Local Similarity 24.6%; Pred. No. 0.0021;

Matches 83; Conservative 44; Mismatches 130; Indels 80; Gaps 19;

QY 101 VEYQWT-----FGGVPQGMIRVREGDTIEVQFSN--HPDSKPHNVDFHATG 148  
DB 40 VEYMWADPCNENIVGINGQFPPTIRANAGDTIVVELINKLHTEGVVH--WHGILQ 96

QY 149 PG-----GGAASFTA--FGHTSTFSFKALQPLGVYVHCAPVGMHIANGMYGLIIVEP 201  
DB 97 RGTWADGTASISQCAINPGEFFNFVNDNGTFYH--GHLMQSGAGLYGLIYDP 153

QY 202 KEGLPK--VDKEYVMQGDFTYTKGKYGQ-GLQPFDEKAIK-EDAEVYVNG----- 250  
DB 154 PGGKPEPHYDGEINLLISDWCHQSINHOEVLG-----SKPIRWIGEPQTLILNGRQFD 209

QY 251 -SVGALTGEN--ALKAKVGETVRLFYVNGGPNLTSSFFHVIIGEIFDKVHREGKGNNHQ 307  
DB 210 CSIAKYDSTNLEPCKLKGSEPCAPY-----FVMPKTYRINI-----AS 250

QY 308 TTLIPAGGAATTE---FVYDVGQDY---LVDAIFPAFNKALGILKVEGENHEIY- 359  
DB 251 TTLAALANPAINHNLVLEADGVNYGFFYSIDIDYSGESVSLITDQNSSENYWVSV 310

QY 360 ---SHKQTD---AVLPEGAPOAIDTQEAAPT 387  
DB 311 GTRGRHPMTPPGLTLNLYLPNSVSK-LPTSPPEPTPA 346

# RESULT 28

S66353 L-ascorbate oxidase (EC 1.10.3.3) precursor - common tobacco

C.Species: Nicotiana tabacum (common tobacco)

C.Date: 28-Oct-1996 #sequence\_revision 13-Mar-1997 #text\_change 20-Jun-2000

C.Accession: S66353

R.Kato, N.; Esaka, M.

Plant Mol. Biol. 30, 833-837, 1996

A.Title: cDNA cloning and gene expression of ascorbate oxidase in tobacco.

A.Reference number: S66353; MUID:96194464; PMID:8624413

A.Accession: S66353

A.Status: preliminary; nucleic acid sequence not shown

A.Molecule type: mRNA

A.Residues: 1-578 <KAT>

A.Cross-References: EMBL:D43624; NID:g599593; PIDN:BA07734.1; PID:g599594

C.Superfamily: laccase

C.Keywords: oxidoreductase

F.373-565/Domain: carboxyl-terminal beta-barrel #status predicted <BB3>

F.87-475/Binding site: copper (His) (type 2) #status predicted

F.89,131,133,477,533,535/Binding site: 2Cu-O cluster (His) (copper type 3) #status predi

F.108-565/Disulfide bonds: #status predicted

F.190,390,470,539/Binding site: substrate (Trp, Trp, Glu, His) #status predicted

F.207-221/Disulfide bonds: #status predicted

Query Match 6.2%; Score 162; DB 2; Length 578;

Best Local Similarity 22.9%; Pred. No. 0.0025;

Matches 104; Conservative 55; Mismatches 162; Indels 134; Gaps 23;

QY 91 VEKWRMLADGVEYQFWTGGQVPGQMIRVREGDTIEVQFSN--HPDSKPHNVDFHATG 148  
DB 37 VEYIHWSPDEGSVVMGINGQFPPTIRAKGDTVAVHLTKLHTEGVVH--WHGIRQ 93

QY 149 PG-----GGAASFTA--FGHTSTFSFKALQPLGVYVHCAPVGMHIANGMYGLIIVEP 201  
DB 94 IGTWADGTASISQCAINPGETFLRFVNDAGTYFH--GHYMQSGAGLYGLIYEV 150

QY 202 KEGLPK--VDKEYVMQGDFTYTKGKYGQGLQPFDEKAIKREDAEVYVNGSGALTGE 258  
DB 151 GEGKEPEPHYDGEFLLISDWMHKGSHE---QEVDL-----SS 185

QY 259 NALKAKVGETVRLFY-VNGGPNLTSSFFHVIIGEIFDKVHREGKGNNHI----- 306  
DB 186 NPLR-WIGEPQTLILNGRQYNCISLAFSPKPLPQCKLRGBOYAPQILRVNPKIYRL 244

QY 307 ---OTTLPAGGAATTEK---VDVPGDYVLVDHAIFFAFNKGALGILKVEGENHEIYS 360  
DB 245 RVASTLNGSLISLAIIGKHWVLEADGVY-----QPFs-----VDMDIYS 286

QY 361 HKQTDVAVLPEGAPO---AIDTQ-EAPKTPAPANIQEOIKAGKATYDSCAACHPDG 414  
DB 287 GESYSVLEKTDQDPTKNWISINVRGEPKTPQGLTL-----LNYLPNSAS----- 332

QY 415 KGVNAPFPLAN--SDYLADHARASIVANGLSGK-----ITVNG----- 453  
 DB 333 -KFTLPPLAPLAWNDY---NHSKSPNKIPALMGSPKPPPPONHRIITLNTONKIDYGT 388  
 QY 454 ---NQESVMPALALSDQOIANVITYTLNSFGNK 484  
 DB 389 KKAINVSLVLPF-----QVYLSINXGINAFDTK 418

## RESULT 29

CCTM5T

cytochrome c552 [validated] - Thermus aquaticus

C:Species: Thermus aquaticus

C:Date: 04-Dec-1986 #sequence\_revision 04-Dec-1986 #text\_change 17-Nov-2000

C:Accession: A00112

R:Titani, K.; Ericsson, L.H.; Hon-nami, K.; Miyazawa, T.

B:Biochem. Biophys. Res. Commun. 128, 781-787, 1985

A:Title: Amino acid sequence of cytochrome c-552 from Thermus thermophilus HB8.

A:Reference number: A00112; PMID:85199131; PMID:2986626

A:Accession: A00112

A:Molecule type: protein

A:Residues: 1-131 &lt;TIT&gt;

A:Note: the source was designated as Thermus thermophilus

R:Tham, M.E.; Hof, P.; Huber, R.; Bourenkov, G.P.; Bartunik, H.D.; Buse, G.; Soulimane,

submitted to the Protein Data Bank, June 1997

A:Reference number: A77356; PDB:1C52

A:Contents: annotation; X-ray crystallography, 1.28 angstroms, residues 1-131

R:Tham, M.E.; Hof, P.; Huber, R.; Bourenkov, G.P.; Bartunik, H.D.; Buse, G.; Soulimane,

J. Mol. Biol. 271, 629-644, 1997

A:Title: Thermus thermophilus cytochrome c-552: a new highly thermostable cytochrome-c sd

A:Reference number: A59160; PMID:97428333; PMID:9281430

A:Contents: annotation; X-ray crystallography, 1.28 angstroms

C:Comment: This cytochrome appears to function as an electron donor to cytochrome oxidase

C:Superfamily: T. aquaticus cytochrome c552; cytochrome c6 homology

C:Keywords: cytochrome; electron transfer; heme; iron; metalloprotein; pyroglyutamic acid

F:2-87/Domain: cytochrome c6 homology &lt;CY6&gt;

F:1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental

F:11/Binding site: heme (Cys) (covalent) #status experimental

F:15,69/Binding site: heme iron (His, Met) (axial ligands) #status experimental

Query Match 6.1%; Score 161; DB 1; Length 131;

Best Local Similarity 37.3%; Pred. No. 0.00042;

Matches 44; Conservative 23; Mismatches 39; Indels 12; Gaps 6;

QY 396 KAGKATYDNCACHPDGKGVNAPPLAN--SDYLADHARASIVA--NGISGKITV 451

DB 1 QADGAKIYAOCACGCHQONGGOGIPGAPPLAGHVAELIAKEGSEYLLVLYGLGQGIYV 60

QY 452 NGNOYSEVMPALALSDQOIANVITYTLNSFGN-----KGGQ-LSADLVAK--AKTKP 501

DB 61 KGMKYNQVMSFPALQKDEIAPVLAHLITATMGDAKKYKGFPTAEVKKLRKKLTP 118

## RESULT 30

C82845

copper resistance protein A precursor XFO132 [imported] - Xylella fastidiosa (strain 9a5

C:Species: Xylella fastidiosa

C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 17-Nov-2000

C:Accession: C82845

R:Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequat

Nature 406, 151-157, 2000

A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.

A:Reference number: A82515; PMID:20365717; PMID:10910347

A:Note: for a complete list of authors see reference number A59328 below

A:Accession: C82845

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-611 &lt;SIM&gt;

A:Cross-references: GB:AE003866; GB:AE003849; NID:99104906; PIDN:AAF82945.1; GSPDB:GN001

R:Simpon, A.J.G.; Reinach, F.C.; Artuda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A

Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carver, H

as-Neto, E.; Docena, C.; El-Dorri, H.; Facincanti, A.P.; Ferreira, A.J.S.

Submitted to GenBank, June 2000

A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; From

J.D.; Unquerra, M.L.; Kemper, E.L.; Kitchima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigr

Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E

A:Authors: Martins, E.M.F.; Matukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;

F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmetti, D.A

Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak

A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silvair

M.; Teunako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z

A:Reference number: A59328

A:Contents: annotation

C:Genetics:

A:Gene: XFO132

C:Superfamily: laccase

Query Match 6.1%; Score 159.5; DB 2; Length 611;

Best Local Similarity 20.1%; Pred. No. 0.004;

Matches 95; Conservative 57; Mismatches 136; Indels 185; Gaps 22;

QY 59 LEVIDAIVTHA--PEVPPVDRDHAKVYVVKMETEKVRLADGVEYGFRTGGGVPGOM 116

DB 39 LEV-----HAAPLPLPQRTTLELRLIGSSVNFGRKMPAL-----TVNLSLPGV 87

QY 117 IFRREGDTIEVQFSNH-PDSKMPHNVDPHAATGPGG-----GAESAFTAPGHTSTSPFKA 170

DB 88 LEMRGDPTVQIHVNTLPD--VMTSIHMHGVLPSNMGGVPGMSPDGIAFGSHLYLRFOL 145

QY 171 LOPGLYTHCAVAPVGMHIANMGYLLVPEKGLP-KYDKYVYMGDDFTYKKGKYSBG 229

DB 146 HOSGTWYH---SHMPQEOAGLYGALIIDPLEPPYRADREHILFSDW-----TD 194

QY 230 LOPPEMEKIRBDAYVNGSVGALTEGNALAKAVGEVRFVYNGGPNLTSSPFVIGE 289

DB 195 LPPALFRLKMKSSI-----DNTYQ-----RTVDF-----FHDIT-- 225

QY 290 IFDKVFEKGKGNHNIQTLL-----IPAG----- 314

DB 226 -----RRDGLRTLLADRMGMGRMTPTDLSVNAHTYTYLLNGTTPAGAWTG 273

QY 315 -----GAITEPKVDVPGDYLVVDHAIFFAFKAGALIKVGEENHEIYS 360

DB 274 LRPGEKVLRLINGSAMTYFVDIRPG-----LKLTVAVDQ-----YV 313

QY 361 HKQT-----DANYLPEGAPO-AIDTQBAKT-----P 386

DB 314 HPTVDELRIALAEYVDVLIQPHGDAFAIPADMGRTYAGCTLAVRGLAPRLALDP 373

QY 387 APANLQEQIKAGKATYDNCACHPDGKGVNAPPLANSDYLNADHARAS 439

DB 374 RRLTWMQDMGHMAGHGHSRASMD-----MPCMTHPRL-GDHASHAS 418

## RESULT 31

F70813

hypothetical protein RY0846c - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000

C:Accession: F70813

R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S

; Connor, R.; Davies, R.; Devlin, K.; Feldwell, T.; Gentles, S.; Hamlin, N.; Holtroyd, S

Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A:Authors: squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrall, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; PMID:98295987; PMID:9634230

A:Accession: F70813

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1504 &lt;COL&gt;

A:Cross-references: GB:AL022004; GB:AL123456; NID:93261550; PIDN:CAI17652.1; PID:929169C

A:Experimental source: strain H37RV

C:Genetics:

A:Gene: Rv0846c  
C:Superfamily: laccase

Query Match 6.0%; Score 158.5; DB 2; Length 504;  
Best Local Similarity 20.8%; Pred. No. 0.0036;  
Matches 104; Conservative 51; Mismatches 180; Indels 165; Gaps 23;

QY 16 SALMLSGCSNQADKAAQPSSTVDAATAKTANADNA-ASQEHQGLPVIDAIVTHAEVPP 74  
DB SGPALAAC-----ASKPTAS--GAAGMTAIDAEEARPHSR--TVATATL-----Pq 72  
QY 75 PVDROHPAYVVMKEVEKMLADGVEYQFTFGGQVPGQMRVREBGTIEVQFSNHPD 134  
DB PARIDGCPVSTL-----TYGNTIPGLPLRAVGVGEIVVSVNRLG 114  
QY 135 SKMHPNVDFH--AATPGGGAASFT--ARGHTSPFKALQGLVYVHCAPVGMHTA 190  
DB 115 D--PTSVHMHGIALRNDMDSTEPATNIGPGDFTFRFSVPDPGTWAPHVGLQGDH-- 170  
QY 191 NGMYGLILVEPKKGLPKVDKEYVVMQDFTYTKG-----KYGEQGLQPFDEKAIRED 243  
DB 171 -GLYLVVVVDDPTPEPHYAEWIIILD--WTDSIGKSPQOLYGL-----TDPKPKPTMQT 224  
QY 244 -----EYVFNCSVCLTGENALKAKVGETVRLFVNGGPNL 280  
DB 225 TGMPEGEVDVSNLLGGDGDIAVPYLLNGRIP--VAATSFKAKPGQIRIRIINSAADT 282  
QY 281 TTSFVHIGEIFDKVHPEG-----GKGENHNIQTLLIPAGG-----AATPEFY 323  
DB 283 ARRIALAGSMVTYTHDGPVIFTEVDALLIGAEIRYDWT--AAGGVFPLVALLEGK- 339  
QY 324 DVPGDVYLVDAIFRAFNKALGILKVEGENHEIYSHKOTDAVYLPFGAPQADTQEAR 383  
DB 340 -----MLARALLSTGAG-----SPDDQFRPDELNMRGVGVNF 374  
QY 384 KTRPAPNLOEQIKAGATYDSNCACHQDPGKGVNAPPLANSDYLDNDHAPASIVAN 443  
DB 375 TAATTANL-----GRPEPHDLP-----VTLG 396  
QY 444 GLSGKI--TVNGNYESVWP 461  
DB 397 GTMAKIDWTINGEPISTNTP 416

## RESULT 32

T35030  
probable copper oxidase - Streptomyces coelicolor  
C:Species: Streptomyces coelicolor  
C:Date: 05-Nov-1999 #sequence\_revision 05-Nov-1999 #text\_change 05-Nov-1999  
C:Accession: T35030

R:Seeger, S.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.  
submitted to the EMBL Data Library, June 1999  
A:Reference number: Z21565  
A:Accession: T35030  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-343 <SEE>  
A:Cross-references: EMBL:AL079355, PTDN:CA845586.1; GSPDB:GN00070; SCOEDB:SC4C6.22  
A:Experimental source: strain A3(2)  
C:Genetics:  
A:Gene: SCOEDB:SC4C6.22

Query Match 5.9%; Score 154; DB 2; Length 343;  
Best Local Similarity 22.0%; Pred. No. 0.0044;  
Matches 79; Conservative 48; Mismatches 114; Indels 118; Gaps 17;

QY 96 RLADG-VEYQFTFGQVPGQMRVREBGTIEVQFSNHPD--SKMHPNVDFHAATPGG 151  
DB 57 KLADGQMGVGFEGKASVPGPLIEVNEGTLHIEFTNTDVRASLHVHGLDYEISS--DGT 115  
QY 152 GABAGFTAGHTSTSFKALQP-----GLVYVHCAPVGMH-----IANGM 193  
DB 116 AMKSDVEEGGTTRTYMTRHKPGRDDGTWRPGSAGYWHYHDV--VGTEHGTGIRNGL 173

QY 194 YGLILVEPKKGLPKVDKEYVVMQDFTYTKGYEGQGLQPFDEKAIREDAEY-VVFNCSV 252  
DB 174 YGPVIVRR-----GD-----VLPAITHIVND-- 197  
QY 253 GALTGENA-----LKAVGETVRLFVNGGPNLTSFHYIGELFDKVFEGKGENIN 305  
DB 198 --MTINNRKPHGTDPFEATVGDREIVM-----LTH-----GEYHTFHMHGHRWADR 244  
QY 306 IQTTLIPAGGAITEFKVDVPGDYLVDHAIIRAFNKGALGILKVEG-----ENNE 357  
DB 245 TGLIGPDDPSRVITDNKLTGPAD-----SFGQIITAGEGVGAGAMMYHCH 289  
QY 358 IYSHKOTDAVYI-----PEGAPQADTQEARPTAPANLQEQIKAGATYDSNCACHQ 411  
DB 290 VQSHSDMGVNGVLFYKKPQGTIPGVREPHHGATA-----KSGESEPFGGAANE 340

## RESULT 33

KSKYAO

L-ascorbate oxidase (EC 1.10.3.3) precursor - cucumber

N:Alternate names: ascorbase

C:Species: Cucumis sativus (cucumber)

C:Date: 30-Sep-1991 #sequence\_revision 30-Sep-1991 #text\_change 11-Jun-1999

C:Accession: A30094

R:Ohkawa, J.; Okada, N.; Shimmyo, A.; Takano, M.  
Proc. Natl. Acad. Sci. U.S.A. 86, 1239-1243, 1989

A:Title: Primary structure of cucumber (Cucumis sativus) ascorbate oxidase deduced from

A:Reference number: A30094; MOID:89145218; PMID:2919172

A:Accession: A30094

A:Molecule type: mRNA

A:Residues: 1-587 <OHK>

A:Cross-references: GB:J04494; NID:9167512; PTDN:AAA33119.1; PID:9167513

A:Comment: This enzyme, which catalyzes the oxidation of L-ascorbate to dehydroascorbate

C:Superfamily: laccase

C:Keywords: copper; glycoprotein; oxidoreductase

F:1-33/Domain: signal sequence #status predicted <SIG>

F:34-587/Product: L-ascorbate oxidase #status predicted <MAT>

F:38-168/Domain: amino-terminal beta-barrel #status predicted <BB1>

F:169-346/Domain: middle beta-barrel #status predicted <BB2>

F:379-574/Domain: carboxyl-terminal beta-barrel #status predicted <BB3>

F:54-236, 116-574, 215-228/Disulfide bonds: #status predicted

F:95,483/Binding site: copper (His) (type 2) #status predicted

F:97,139, 141,485,542,544/Binding site: 2Cu-O cluster (His) (copper type 3) #status predi

F:360,401,475/Binding site: carbohydrate (Asn) (covalent) #status predicted

F:480,543,548,553/Binding site: copper (His, Cys, His, Met) (type 1) #status predicted

Query Match 5.8%; Score 153.5; DB 1; Length 587;  
Best Local Similarity 28.0%; Pred. No. 0.0097;  
Matches 49; Conservative 91; Mismatches 60; Indels 35; Gaps 10;

QY 101 VEYQFT-----FGQVPGQMRVREBGTIEVQFSN--HPDSKMPHNVDFHAATG 148  
DB 45 VEYFMSPDCVENIWMGIVNGEPGPTIRANAGDIVVELTNLHTEGVVH--WHGIQ 101  
QY 149 PG-----GGAASFTA--PGHTSPFKALQGLVYVHCAPVGMHTANGMYGLIYVP 201  
DB 102 RCTPMADGTASISQCAINPGETTFRFVYDKAGTYFH--CHLGQNSAGLIGSLIYDP 158  
QY 202 KEGLPK--VDKEYVVMQDFTYTKGYEQ--GLQPFDEKAI--EDAEVVFNG 250  
DB 159 PEGRSEPHYDEINLLSDMWHQSVHKEVGLS-----SKPMRWIGEPOSILING 209

## RESULT 34

A83363

probable cytochrome c precursor PA2266 (imported) - Pseudomonas aeruginosa (strain PA01)

C:Species: Pseudomonas aeruginosa

C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000

C:Accession: A83363

R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Br  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Laidig, K.; Lam,  
., Lozy, S.; Olson, M.V.



Nature 406, 959-964, 2000  
A>Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen  
A/Reference number: AB2950; MUID:20437337; PMID:10984043  
A/Accession: AB3363  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-439 <STO>  
A/Cross-references: GB:AE004657; GB:AE004091; NID:99948287; PIDN:AA05654.1; GSPDB:GN001  
A/Experimental source: strain PA01  
A/Genetic: P2266  
A/Gene: PA2266  
C/Superfamily: membrane-bound alcohol dehydrogenase cytochrome c; cytochrome c6 homology  
C/Keywords: chromoprotein; heme; iron; metalloprotein  
F/40/43/Binding site: heme (Cys) (covalent) #status predicted  
F/44/Binding site: heme iron (His) (axial ligand) #status predicted  
F/186/191/Binding site: heme (Cys) (covalent) #status predicted  
F/192/Binding site: heme iron (His) (axial ligand) #status predicted  
F/324/327/Binding site: heme (Cys) (covalent) #status predicted  
F/328/Binding site: heme iron (His) (axial ligand) #status predicted

Query Match 5.8%; Score 153; DB 2; Length 439;  
Best Local Similarity 35.5%; Pred. No. 0.0071;  
Matches 38; Conservative 15; Mismatches 48; Indels 6; Gaps 4;

Db 396 KAGKATYDSNCAACHOPDGKGVNPAFPPLA-NSDYLNADHARAASIVANGLSKITVNGN 454  
314 KTGAALYVDNCGACHRDGKGYARVPALAGNPTVTSPTSLSVHVLKG-GLTPATHQ 371

Qy 455 QYES-VMPALA--LSDQIANVITYTLNFSGNKGQSLADVAKAKK 498  
372 APSSTFMPFGMRNNDDEIDAVVNFIRTSWGNQSPSVDEVRLRK 418

RESULT 35  
T04343  
L-ascorbate oxidase (EC 1.10.3.3) - rice (fragment)  
C/Species: *Oryza sativa* (rice)  
C/Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 20-Jun-2000  
C/Accession: T04343  
R/Tabata, K.; Hirose, A.; Saaka, M.  
Submitted to the EMBL Data Library, June 1997  
A/Description: Gene expression and function of ascorbate oxidase in rice.  
A/Reference number: 215298  
A/Accession: T04343  
A/Status: translated from GB/EMBL/DBJ  
A/Molecule type: mRNA  
A/Residues: 1-380 <TAB>  
A/Cross-references: EMBL:AB004799; PIDN:BA20520.1  
A/Experimental source: subsp. *Gramineae*  
C/Superfamily: laccase  
C/Keywords: oxidoreductase  
F/21-237/Disulfide bonds: #status predicted

Query Match 5.8%; Score 151.5; DB 2; Length 380;  
Best Local Similarity 22.4%; Pred. No. 0.0075;  
Matches 88; Conservative 47; Mismatches 138; Indels 119; Gaps 17;

Qy 101 VEQFWT-----FGQVPGOMTRVREGDTIEVQFSN--HPDSKPHNVDFNAATG 148  
12 VEVYLMAPDCQORVMIGINGRFPGRNITARAGDVISVTMNNKMTGCVIHMGIROFGT 71  
Db 149 PGGAASFTA-----PGHTSFSPKALOPGLYVHCANAVPCHMIANGMGLIVERKEG 204  
72 PMADGTASISOCANVPSTFYKFDVADPFTYH--GHFGMRRAAGLIGSLIVLDSPE 128

Qy 205 LPKVDKEY-----VMQDFTYKGYG-QLQCFDMEKAIREDAEVVFNG----- 250  
Db 129 QPEFRHQYDDGGLPMMLSLDMHNVVYAGALDCKDHNFEWIGRPQIILNGRQFE 188  
Qy 251 -SYG-----ALTGENALAKY-----GE 267  
Db 189 CTIGPARKSEKLNQERGPASTIRRCATRRSAREERVGRCYPRSGCAPVVFVNWGK 248

Qy 268 TVRLFVNGGPNLTSSFHVIGELIPDKVHFEKGGENHNIQTLLPAGAAITFEKVDVPG 327  
Db 249 TYRLALIAS-----TSLSLT-----NVKIQGNK-----MTVADGNHVEPFVDDID 291

Qy 328 DYVLVDHAI-FAFNKALGILKVEGENH-----EYSHKQTD-----VYLPFGAPQ 375  
Db 292 IYSGSYSVFFKADCKPASWISVGRGHPKTVPALAILSYGNMAAPPLQLPAGEP- 350

Qy 376 AIDTQEAPTKPAALQE-----QIKAGKAT 401  
Db 351 -----PVTAMNDTORSKAFITYSIRAKDT 375

RESULT 36  
F83631  
cytochrome c oxidase, subunit II PA0105 [imported] - *Pseudomonas aeruginosa* (strain PA01)  
C/Species: *Pseudomonas aeruginosa*  
C/Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C/Accession: F83631  
R/Stover, C.K.; Pham, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Bradman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kae, A.; Lardig, K.; Lam, J.; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A>Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen  
A/Reference number: AB2950; MUID:20437337; PMID:10984043  
A/Accession: F83631  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-374 <STO>  
A/Cross-references: GB:AE004449; GB:AE004091; NID:99945928; PIDN:AA03495.1; GSPDB:GN001  
A/Experimental source: strain PA01  
C/Genetic: COXB; PA0105

Query Match 5.7%; Score 150; DB 2; Length 374;  
Best Local Similarity 22.0%; Pred. No. 0.0093;  
Matches 63; Conservative 38; Mismatches 86; Indels 100; Gaps 11;

Qy 265 VGETVRLFVNGGPNLTSSFHVIGELIPDKVHFEKGGENHNIQT--TLIPAGA----- 316  
Db 135 LQGDVEYFENLNPQ-----DOIHNRQAKDEHYLVEBPLVLPVGTQVRLPI 182

Qy 317 -----ATTEF-----KYDVPQDY-----VLVDHA 335  
Db 183 TSSDVIHSMWVPAFAVKRDAIPGFVNEAMTKVDEIGIYRQCAELCGKHGFMPIVD-- 240

Qy 336 IFPAENKALGILKVEGENHETIYSHKQTDAYLDEGAPQALDTEAPKTPAPALQEOI 395  
Db 241 -----VKPKAEFDQWLAKRKE-EAAKVKR-----LTSKEMTK-----EBLV 275

Qy 396 KAGKATYDSNCAACHOPDGKGVNPAFPPLANSVDYLNADHARAASIVANGLSKITVNGN 455  
Db 276 ARGDKYHTITCAACHQABGQGMPEFPAKSGKITVGPKEHLEVVFNQVPG----- 327

Qy 456 YESVMPALA--LSDQIANVITYTLNFSGNKGQSLADVAKAKTK 500  
Db 328 --TAMAFQKQLEVDLAIVITYERNAMGNDGDWTPRDVAVAYQK 372

RESULT 37  
E83075  
probable cytochrome c PA4571 [imported] - *Pseudomonas aeruginosa* (strain PA01)  
C/Species: *Pseudomonas aeruginosa*  
C/Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C/Accession: E83075  
R/Stover, C.K.; Pham, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Bradman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kae, A.; Lardig, K.; Lam, J.; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A>Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen  
A/Reference number: AB2950; MUID:20437337; PMID:10984043  
A/Accession: E83075  
A/Status: preliminary



A.Molecule type: DNA  
A.Residues: 1-675 <STO>  
A.Cross-references: GB:AE004870; GB:AE004091; NID:g9950807; PIDN:AAG07959.1; GSPDB:GN001  
A.Experimental source: strain PA01  
C.Genetics:  
A.Gene: PA4571

Query Match 5.7%; Score 150; DB 2; Length 675;  
Best Local Similarity 26.3%; Pred. No. 0.02;  
Matches 46; Conservative 28; Mismatches 73; Indels 28; Gaps 6;

QY ILKVGSENHETYSKQTDVAV-----LPGAPQAITQEAAPKTP 386  
DB 244 LKRT-GQRFATVSGSMTDVHHSIQFSDDDLALASYLKSIPACKDLPMPDSERPPLA 302  
QY 387 APANTQEQIKAGKATYDSNCAACHOPDGKGVNAPPPILANSYDLNADHARAASIVANGLS 446  
DB 303 APVDLYS--SRGGLGYAQCSCDCKKDSGVGMFPPLGNPTVAS--ANPSTLHITLT 358  
QY 447 GKTVNGNQYESV--MPALA-LSDOQIANVITYTILNSFGNKGQSLSADDVAKAK 498  
DB 359 GVKTAQTATSHSVYTMPGFAQLEDEIREIELISFVSSSWGNGSSSIDAGGVKKLRQ 413

RESULT 38  
F83387  
copper resistance protein A precursor PA2065 [imported] - Pseudomonas aeruginosa (strain C)Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #ext\_change 31-Dec-2000  
C.Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #ext\_change 31-Dec-2000  
C.Accession: F83387  
R.Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Bradman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lardig, K.; Lim, J.; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A.Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho  
A.Reference number: A82950; MUID:20437337; PMID:10984043  
A.Accession: F83387  
A.Status: preliminary  
A.Molecule type: DNA  
A.Residues: 1-632 <STO>  
A.Cross-references: GB:AE004633; GB:AE004091; NID:g9948070; PIDN:AA05453.1; GSPDB:GN001  
A.Experimental source: strain PA01  
C.Genetics:  
A.Gene: pcoa; PA2065  
C.Superfamily: laccase

Query Match 5.7%; Score 149; DB 2; Length 632;  
Best Local Similarity 21.3%; Pred. No. 0.022;  
Matches 87; Conservative 59; Mismatches 149; Indels 114; Gaps 21;

QY 107 TFGGVPGOMIRVREGDTIEVOFSNH--PDSK-----MPHNVDFHATGPGGAEASF 157  
DB 66 TINGSLPPTLRMRGSDVTLAVRRRLAEDTSHHGGIILPANMD-----GYPGLSF 117  
QY 158 --TAGHTSTSFALQGLVYVHCAVAPVGNHIANGMVGLILVEPKGLP-KVDKEYVV 214  
DB 118 EGIAPGGLYEVRFKVRNGTYWYH--SHSGLOEQAGVYALVIDAREPEPSYRDYV 174  
QY 215 MCGPPTYTKGKGEOGLQFPDEKAIREDAEVYV-----NGSICALTGEN- 259  
DB 175 LLSWSDS-----KQRIILAKLKOSDYVNFHRTVGDFTIDVSNANGMATLADRM 226  
QY 260 --ALK-----AKVGETVRLFVNGGP--NLTSFHVIGELFDKVFEGSGEHNHQT 308  
DB 227 WAEKMSPTDLADVSGYTYTYTILNGQPPDGNWTGLFR-----PGEKRLRF 272  
QY 309 TLIPAGAAITEFKVDVPG-----DYVLVDH--AIFRAFNKALGILKVEG 352  
DB 273 ----VNASAMSYFDRIFGLKMTVAADQGHVEPVSDBLRAVAETVDV-----LVEPG 324  
QY 353 EENHEIYSHKQTDVAVLPEGAQALDTQEAAP-KTAPANLQEQIRAKATYDSNC--AA 408  
DB 325 ERAYTLFAQMSDRSGY-ARGTLALAEGLSAPVPTDPRPLIGMDMGMGMGDMGAMGHA 383

QY 409 CHOP-----DGKGVNAPPPILANSYDLNADHARAASIVANGL 445  
DB 384 ARPRPASEMDHSGKMGMDVTGMGDS--KKAQMGMSGMDHSGKMGMDMNGM 430

RESULT 39  
S14271  
membrane-bound alcohol dehydrogenase (EC 1.1.1.-) cytochrome c precursor - Acetobacter i  
N.Alternate names: membrane-bound alcohol dehydrogenase 44k chain  
C.Species: Acetobacter polyoxogenes  
A.Variety: strain NB11028  
C.Date: 21-Nov-1993 #sequence\_revision 09-Aug-1996 #ext\_change 17-Nov-2000  
R.Tamaki, T.; Fukaya, M.; Takemura, H.; Toyama, K.; Okumura, H.; Kawamura, Y.; Nishiyama  
Biochim. Biophys. Acta 1088, 292-300, 1991  
C.Accession: S14271  
A.Title: Cloning and sequencing of the gene cluster encoding two subunits of membrane-bc  
A.Reference number: S14270; MUID:91159482; PMID:2001402  
A.Accession: S14271  
A.Molecule type: DNA  
A.Residues: 1-468 <TAM>  
A.Cross-references: GB:D00635; NID:g216185; PIDN:BA00529.1; PID:g216187  
A.Experimental source: strain NB11028  
C.Complex: heterodimer of 72k and 44k (cytochrome c) chains  
C.Superfamily: membrane-bound alcohol dehydrogenase cytochrome c; cytochrome c6 homology  
C.Keywords: alcohol metabolism; blocked amino end; chromoprotein; electron transfer; hem  
F.1-23/Domains: signal sequence #status predicted <SIG>  
F.320-468/Product: membrane-bound alcohol dehydrogenase cytochrome c chain #status predic  
F.320-403/Domains: cytochrome c6 homology <Cys>  
F.345-48/Binding site: heme (Cys) (covalent) #status predicted  
F.149/Binding site: heme iron (His) (axial ligand) #status predicted  
F.193-196/Binding site: heme (Cys) (covalent) #status predicted  
F.197/Binding site: heme iron (His) (axial ligand) #status predicted  
F.330-333/Binding site: heme (Cys) (covalent) #status predicted  
F.334/Binding site: heme iron (His) (axial ligand) #status predicted

Query Match 5.6%; Score 146; DB 1; Length 468;  
Best Local Similarity 25.2%; Pred. No. 0.023;  
Matches 84; Conservative 50; Mismatches 127; Indels 72; Gaps 19;

QY 184 PYGNHIANGMVGLIL-VEPKGLPVVDKEYVYMGDFYTKKYGQGL---QFPDM-EK 237  
DB 149 PLSMRWPLGIWRMSPSPKCDFTPAFGDPETARGDYLVTPG-GHCGAHTPRGPMQEK 207  
QY 238 AIREDAEVYVNGSGALITGENALAKAYGETYR-----LFVNGGNNLTSFHVIEIPFX 293  
DB 208 AL--DAA-----GGPFLSGAPIDNWAPSLRNDPVVGLGKMSDDIYTFUKSGRIDHS 260  
QY 294 VHEFGKGEHNHNIQTLLIPAGGAITEFKVDVPGDYVLVD--HAIFRAFNKALGILKVE 351  
DB 261 AVF-GGMD-----VVMSTQYFDDDLALAK-YLKSLLPVVPPSQ 299  
QY 352 GEENHEIYSHKQTDVAVLPEGAQALDTQEAAPKTPAPANLQEQIRAKATYDSNCAACHQ 411  
DB 300 GN-----YTYDPSTANMLASG-----NTASVP-----GADTYKECAICHR 335  
QY 412 PDGKGVNAPPPILANSYDLNADHARA-ASIVANGISGKITVNGNQYESV-MPAL--ALSD 467  
DB 336 NDGGGVAMPPPLANPVVVTENPTSLVNVLAHG--GVLPSNMAPSAVAMGYSLSLA 393  
QY 468 QQIANVITYTILNSFGNKG-QQLSADDVAKAKT 499  
DB 394 QQIADVNVFITSWGKAPGVITADYKLAADT 426

RESULT 40  
B49340  
membrane-bound alcohol dehydrogenase (EC 1.1.1.-) cytochrome c precursor - Acetobacter f  
N.Alternate names: membrane-bound alcohol dehydrogenase 44k chain  
C.Species: Acetobacter pasteurianus  
C.Date: 07-Apr-1994 #sequence\_revision 09-Aug-1996 #ext\_change 17-Nov-2000  
C.Accession: B49340  
R.Takemura, H.; Kondo, K.; Horinouchi, S.; Bepnu, T.

J. Bacteriol. 175, 6857-6866, 1993  
 A:Title: Induction by ethanol of alcohol dehydrogenase activity in *Acetobacter pasteurii*  
 A:Reference number: A49340, MUID:94042848, PMID:8226628  
 A:Accession: B49340  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-472 <TAK>  
 A:Cross-references: GB:D13893; NID:6517067; PIDN:BAA02993.1; PID:9452587  
 A:Experimental source: strain NC11380  
 C:Complex: heterodimer of 72K and 44K (cytochrome c) chains  
 C:Superfamily: membrane-bound alcohol dehydrogenase cytochrome c; cytochrome c6 homology  
 C:Keywords: alcohol metabolism; blocked amino end; chromoprotein; electron transfer; hem  
 F:1-24/Domain: signal sequence #status predicted <SIG>  
 F:25-472/Product: membrane-bound alcohol dehydrogenase cytochrome c chain #status predic  
 F:321-404/Domain: cytochrome c6 homology <CY6>  
 F:46-49/Binding site: heme (Cys) (covalent) #status predicted  
 F:50/Binding site: heme iron (His) (axial ligand) #status predicted  
 F:194,197/Binding site: heme (Cys) (covalent) #status predicted  
 F:198/Binding site: heme iron (His) (axial ligand) #status predicted  
 F:331,334/Binding site: heme (Cys) (covalent) #status predicted  
 F:335/Binding site: heme iron (His) (axial ligand) #status predicted.

## Query Match

5.5%; Score 145.5; DB 1; Length 472;  
 Best Local Similarity 33.3%; Pred. No. 0.025;

Matches 39; Conservative 18; Mismatches 31; Indels 29; Gaps 5;

Qy	397	AGKATYDSNCA	CHOPDGKVPNAPPLANS	DYLNADH-ARASIVANGIS	SKITVNGND	455
Db	322	AGKTYVEOCAL	CHRDGGVARMFPLAGN	PVVSNDPTSV	NAHIVDG-----	370
Qy	456	YESVMP-----	AIA-----	LSDOQIANVIT	TYTLNSFGNG-GOLSAD	VDVAKAK 497
Db	371	--GVLEPTWAPSA	VAMPDYKNILSDQI	ADVNFIRSAWGNRA	PANTPADIQ	KLR 425

Search completed: August 27, 2003, 18:32:44

Job time : 46 secs



QY 336 IFRAFNKALGILKVEGENHEIYSHKOTDAVYLPEGAPOAIDTOEAPKTPAPANLOEQI 395  
 Db 241 -----VKRAEFDQWLAKRKE-EAAKYKE-----LTSKEWTK-----ELV 275  
 QY 396 KAGKATVNSCAACHQDQSGVPAFPPLANSIDLNDHARAASIVANGLSGKITVNGQ 455  
 Db 276 ARGDKYVHTICACHQAGGQMPMPALGSKITVTPKPKHLELVNGLVPG----- 327  
 QY 456 YESVPAIA--LSDQIANYITTYLNSFGNGGQLSADDAVAKAKTK 500  
 Db 328 --TAMAFGKQLINEVDLAAVITTYERNMGNDGDMTPKQVAVAKOK 372

RESULT 2  
 US-09-732-350-5  
 ; Sequence 5, Application US/09732350  
 ; Patent No. US20010031490A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Svendsen, Allan  
 ; APPLICANT: Xu, Feng  
 ; TITLE OF INVENTION: LACCASE MUTANTS  
 ; NUMBER OF SEQUENCES: 10  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: No. US20010031490A10 No. US20010031490A1disk of No. US20010031490A1ch  
 ; STREET: 405 Lexington Avenue  
 ; CITY: New York  
 ; STATE: NY  
 ; COUNTRY: USA  
 ; ZIP: 10174  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette  
 ; COMPUTER: IBM Compatible  
 ; OPERATING SYSTEM: DOS  
 ; SOFTWARE: FastSeq for Windows Version 2.0  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/732.350  
 ; FILING DATE:  
 ; CLASSIFICATION:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 09/032,315  
 ; FILING DATE:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Rozek, Carol  
 ; REGISTRATION NUMBER: 36,993  
 ; REFERENCE/DOCKET NUMBER: 5200.200-US  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 212-867-0123  
 ; TELEFAX: 212-878-9655  
 ; INFORMATION FOR SEQ ID NO: 5:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 529 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-09-732-350-5

Query Match 5.4%; Score 141; DB 9; Length 529;  
 Best Local Similarity 20.4%; Pred. No. 0.00026;  
 Matches 103; Conservative 60; Mismatches 216; Indels 126; Gaps 20;

QY 59 LPVIDAIVTHAPEVPPVDRDHPAKVVK--METVEKWRMLADGVEYQFMTFGGOVPGOM 116  
 Db 8 LPLLAAVST-----PAPAAVAVNYKFDIKNNVAVAPDGFQRSIVSVNGVLPGLT 54  
 QY 117 IRRREGDTIEVOFSNH---PDSKMPHNVDFH---AATGGGGAESAFTAP-----GHTS 164  
 Db 55 ITANKGDTLRINVTQLTDPMSMRATTIHHHGLFOATTADEDGPAPVTCPIAQNLSYTY 114  
 QY 165 TFSFKALQPLGVYHCAVAPVGMHIANGYG-LILVER---KEGLPKVDKXYVWQGDY 219  
 Db 115 EIRPLRG-QGTGMWYH--AHLASQYVDGLRGPLVITDPDHPKSRDYVDVDASTVWMLDWM 170

QY 220 Y-----TKGKXGEGQ---LOPFDMEKAIREDAAEVVNGSVGALTGENALAKAKVETV 269  
 Db 171 YHTPAPVLEKQMFSTNNNTALLSPVDSGLINGKGRYV-----CGPAPVRSIVINVKRGKY 225  
 QY 270 RLFVNGGNGNLTSFHVIGIEIDPKHFBEGKGENHIQTTLLPAGS-----AATFEKVD 324  
 Db 226 RLKVINASAIQSFTFSI-----EGHSL--TVIEADGILHQPPLAVDSFQIY 268  
 QY 325 VPGDVLVVDHAIIFRANKGALGILKVEGENHEIYSHKOTDAVYLPEGAPOAIDTOE--- 381  
 Db 269 AGGRYSVIVYEAQNTAANYIIRAPMTVAGAGTANLIDPTNVFAVLHNEGAPNAEPTTEQGS 328  
 QY 382 -----APKTPAPANLOEQIKAGKATYDS-----NCAACHQDPDGKV 417  
 Db 329 AIGTALVEENLHALINPGAPGSAADVSLNLAIGRSTVDGILRFTFNNIKYEAP----- 383  
 QY 418 PNAFPPLANSIDYLNADHAAASIVANGLSGKITVANGNOYESVPAIALSDQIANYITTY 477  
 Db 384 --SLPPL-----LKITLANNASNDADFTPNEHTIVLPNKVIELNITGADHP 428  
 QY 478 LNSFGNGGQLSADDAVAKAKTKRPN 502  
 Db 429 IHLHG-----VFDLVKSLGGTPN 447

RESULT 3  
 US-09-869-877-5  
 ; Sequence 5, Application US/09869877  
 ; Publication No. US20020192792A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Schneider, Palle  
 ; APPLICANT: Danielson, Steffen  
 ; APPLICANT: Svendsen, Allan  
 ; TITLE OF INVENTION: laccase Mutants  
 ; FILE REFERENCE: 10179.204-US  
 ; CURRENT APPLICATION NUMBER: US/09/869.877  
 ; CURRENT FILING DATE: 2001-07-06  
 ; NUMBER OF SEQ ID NOS: 10  
 ; SOFTWARE: Patentin version 3.1  
 ; SEQ ID NO 5  
 ; LENGTH: 529  
 ; TYPE: PRT  
 ; ORGANISM: Rhizoctonia solani  
 ; US-09-869-877-5

Query Match 5.4%; Score 141; DB 10; Length 529;  
 Best Local Similarity 20.4%; Pred. No. 0.00026;  
 Matches 103; Conservative 60; Mismatches 216; Indels 126; Gaps 20;

QY 59 LPVIDAIVTHAPEVPPVDRDHPAKVVK--METVEKWRMLADGVEYQFMTFGGOVPGOM 116  
 Db 8 LPLLAAVST-----PAPAAVAVNYKFDIKNNVAVAPDGFQRSIVSVNGVLPGLT 54  
 QY 117 IRRREGDTIEVOFSNH---PDSKMPHNVDFH---AATGGGGAESAFTAP-----GHTS 164  
 Db 55 ITANKGDTLRINVTQLTDPMSMRATTIHHHGLFOATTADEDGPAPVTCPIAQNLSYTY 114  
 QY 165 TFSFKALQPLGVYHCAVAPVGMHIANGYG-LILVER---KEGLPKVDKXYVWQGDY 219  
 Db 115 EIRPLRG-QGTGMWYH--AHLASQYVDGLRGPLVITDPDHPKSRDYVDVDASTVWMLDWM 170  
 QY 220 Y-----TKGKXGEGQ---LOPFDMEKAIREDAAEVVNGSVGALTGENALAKAKVETV 269  
 Db 171 YHTPAPVLEKQMFSTNNNTALLSPVDSGLINGKGRYV-----CGPAPVRSIVINVKRGKY 225  
 QY 270 RLFVNGGNGNLTSFHVIGIEIDPKHFBEGKGENHIQTTLLPAGS-----AATFEKVD 324  
 Db 226 RLKVINASAIQSFTFSI-----EGHSL--TVIEADGILHQPPLAVDSFQIY 268  
 QY 325 VPGDVLVVDHAIIFRANKGALGILKVEGENHEIYSHKOTDAVYLPEGAPOAIDTOE--- 381  
 Db 269 AGGRYSVIVYEAQNTAANYIIRAPMTVAGAGTANLIDPTNVFAVLHNEGAPNAEPTTEQGS 328

QY 382 -----AKTPAPANLOEQIATKATYDS-----NCACHQDPQGV 417  
DB 329 AIGTALVEENIHALINPGAFGSGAPADVSLNLAIGRSTVDGIRFTFNINIKYAP----- 383  
QY 418 PNAFPLANSDYLNADHARAASIVANGLSKITVNOQVESVPALALSQOJANITYT 477  
DB 384 --SLPFL-----LKITLANNASNDADFTFNEHTIVLPHNKVILELNTGADHP 428  
QY 478 LNSFGNGQLSADVAKAKTKPN 502  
DB 429 IHLHGH-----VFELVKSIGTFN 447  
RESULT 4  
US-09-738-626-6780  
; Sequence 6780, Application US/09738626  
; Publication No. US20020197605A1  
; GENERAL INFORMATION:  
; APPLICANT: NAKAGAWA, SATOSHI  
; APPLICANT: MIZOGUCHI, HIROSHI  
; APPLICANT: ANDO, SEIKO  
; APPLICANT: HAYASHI, MIKIRO  
; APPLICANT: YOKOI, HARUHIKO  
; APPLICANT: TATEISHI, NAOKO  
; APPLICANT: SENOH, AKIHIRO  
; APPLICANT: IKEDA, MASATO  
; APPLICANT: OZAKI, AKIO  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-125  
; CURRENT APPLICATION NUMBER: US/09/738, 626  
; PRIOR FILING DATE: 2000-12-18  
; PRIOR APPLICATION NUMBER: JP 99/377484  
; PRIOR FILING DATE: 1999-12-16  
; PRIOR APPLICATION NUMBER: JP 00/159162  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: JP 00/280988  
; PRIOR FILING DATE: 2000-08-03  
; NUMBER OF SEQ ID NOS: 7059  
; SOFTWARE: PatentIn ver. 3.0  
; SEQ ID NO 6780  
; LENGTH: 493  
; TYPE: PRT  
; ORGANISM: Corynebacterium glutamicum  
US-09-738-626-6780  
Query Match 5.4%; Score 140.5; DB 10; Length 493;  
Best Local Similarity 19.0%; Pred. No. 0.00026;  
Matches 80; Conservative 57; Mismatches 174; Indels 109; Gaps 15;  
QY 15 LSLMLSGSGNQADKAAQPKSSIVDAAKTANMDNNAQDEHGELEFVIDAIVTHAEVPP 74  
DB 12 LGGGLVAG-----TGAVAACTSDPGPAAS-----APG 38  
QY 75 PVDURDPAKVVVMEETVEKVMRL-----ADGVEYQFTFGGQVPGOMIRVREGDTIEVO 128  
DB 39 PSIRPRTPTALCEPFTVRRITLARPISLDIGIEATKMGVSTGDAIEATGVDVQVD 98  
QY 129 FSNH-PDSKMP--HNVDFAATGPGGGAESAFTAPGHTSTFSFKALQPGIYVYHCAVAP 185  
DB 99 ITNELPESTSIHWHGIALHNAADGVPGMTQDPIEPGESFSEYVEVPHGTFYFH---SHT 155  
QY 186 GMIHANGMGLIVERKEGLPKVDKEYVYMGDFYTKGKYGEGLOPDMEK----- 237  
DB 156 GLOLDGLHAPLIRPDQAEQDQDVEWTVLDDWVD---GIQGTPEDELDKLTGMSGSD 211  
QY 238 -----AIRBDAEYVFNESYGAL-----TGENALKAKVGETVRL-FV 273  
DB 212 HNGRMGMGHHGQMMHGTTPRVLGSDGVVYHYLLNGRIAPRHRTFEAPRPGDKALRPI 271  
QY 274 GNGGPNLTSSPHVI--GEIFDKYHFGGKGGENNIOTLLIPAGGAATEPKVDV--GD 328  
DB 272 NSGGDTI---FKVALGCHMTVTHTDGFVQPMETESIYLSMGE-----RVDEVILGD 322

QY 329 YVLVDHAFRANFGALGILKVEGENHE-----IYSHKQTDVAVLEPGAP 374  
DB 323 GIEPLTALVAGDDRAFVAFIRTAGGQAPRPVDOPFELSTGLLSLTKRADRALPEGRP 382  
RESULT 5  
US-10-174-693-291  
; Sequence 291, Application US/10174693  
; Publication No. US20030131373A1  
; GENERAL INFORMATION:  
; APPLICANT: Bloksberg, Leonard N.  
; APPLICANT: Havukkala, Ilkka  
; TITLE OF INVENTION: Materials and Methods for the  
; FILE REFERENCE: 11000.1003c5  
; CURRENT APPLICATION NUMBER: US/10/174, 693  
; PRIOR FILING DATE: 2002-06-18  
; PRIOR APPLICATION NUMBER: US 08/975,316  
; PRIOR FILING DATE: 1997-11-21  
; PRIOR APPLICATION NUMBER: US 08/713,000  
; PRIOR FILING DATE: 1996-09-11  
; PRIOR APPLICATION NUMBER: US 09/169,789  
; PRIOR FILING DATE: 1998-10-09  
; PRIOR APPLICATION NUMBER: US 09/615,192  
; PRIOR FILING DATE: 2000-07-12  
; NUMBER OF SEQ ID NOS: 407  
; SOFTWARE: PaateSeq for Windows Version 3.0  
; SEQ ID NO 291  
; LENGTH: 326  
; TYPE: PRT  
; ORGANISM: Eucalyptus grandis  
US-10-174-693-291  
Query Match 5.2%; Score 136.5; DB 16; Length 326;  
Best Local Similarity 20.9%; Pred. No. 0.00032;  
Matches 71; Conservative 46; Mismatches 129; Indels 93; Gaps 14;  
QY 78 RDHPAKVVVMEETVEKVMRLADGVEYQFTFGGQVPGOMIRVREGDTIEVQPSNHDSDK 137  
DB 9 RNYTFENVKNTT-----RLCS--SKPIYVAGMFPGLTVABEDTVLVRVN---RV 57  
QY 138 PHNVDFH-----AATGPGGAESAFTAPGHTSTFSFKAL-QPGIYVYHCAVAPV 186  
DB 58 KYNVTIHHMGIROLKTMADGPAIYTCPI-QPGSYVYNFTITGKGLLHMAH----- 112  
QY 187 KHIANGMGLIVERKEGL---PVVDKEYVYMGDFYTKGKYG-----EGLQOPDM 236  
DB 113 LMLRATLHGAIYILPRGVYPPPKPKKEVVVVLGSMWKSDEGVISQAISGLAP----- 168  
QY 237 KAIREDAEYVNGSVGALT-----GENALKAKVGETVRLFPVNGGPNLTSSPHVIGEIF 291  
DB 169 ---NVSDHTINGHPGSSNCPGSGFTLLPVESGKYLRIINALNEELFFKLAGH--- 222  
QY 292 DKVHEFGKGGENHNIOTLLIPAGGAATEPKVDVPGDVLVDHAFRANFGALGILKVE 351  
DB 223 -----QLTIVEADATYVAPFKTDT-----IVIAP 246  
QY 352 GEENHEIYSHKQTDVAVLEPG-----APQALDIOEAPKT 385  
DB 247 GQTTNALISTDQSSGKVVVAASPFDSPFAVDNMTATAT 285  
RESULT 6  
US-09-732-350-8  
; Sequence 8, Application US/09732350  
; Patent No. US20010031490A1  
; GENERAL INFORMATION:  
; APPLICANT: Svendsen, Allan  
; APPLICANT: Xu, Feng  
; TITLE OF INVENTION: LACCASE MUTANTS  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:



```

; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738, 626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 4553
; LENGTH: 511
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
; US-09-738-626-4553

```

```

Query Match      5.0%; Score 131.5; DB 10; Length 511;
Best Local Similarity 22.5%; Pred. No. 0.0019;
Matches 104; Conservative 51; Mismatches 199; Indels 109; Gaps 21;

```

```

QY 29 KAAQPKSTVDAAKAANNDNMAOEHQGE---LPVIDAIVTHAPVPPVD--RDHPAK 83
DB 11 KAGVLAATVAGNOVLVACSSDDVRGEGEPTLPY-----PRADGTREGSS 58
QY 84 VVVKMETVEKVRKLADVEYQFTFGGOVPGMIRREGDTIEVQSNHPDS----- 135
DB 59 VHPALAEQGESQILDPDVTTKTGFNGTHLGPVLVVKGDVADVINDENTVTHMG 118
QY 136 -KMPHNVD--HAATPGGGAESFTAPGHTSTFSKALQPGLYVHCAYAVGMIANG 192
DB 119 MTLPALADGPHSPICPGQTWSPWTVANDATLWHPHTHG-----TGLHAYNG 169
QY 193 MYGLILVERKEGLPKVD--KEYVV-----MGDDTYTKGKEGGLQPPDMKAIRED 243
DB 170 LKGMITVE--DEATDKDLPERYGVDDIPLVIMDRLEPDSLDEBPL--DGLLGDTP 226
QY 244 EYVVENGVGALTGENALAKAVGETVRLFYNGGPNL-----TSSFHVIG-----EIPD 292
DB 227 ANGITAHFDATRRVRFRVLNNSNRFY-----NLAFSDRTFQVILASDGLDEPQD 280
QY 293 KHFBSGCKENNNIQTLLIPAGGAIT-----EFKVDVPGDYVLVDHAI FRAFNKALG 346
DB 281 RTLLAIGPERWEIVLELP--GEDVTLSEVGFEDVGVDPDFDPCGSDSFG----- 333
QY 347 ILKVESENHETYSKQTDAAVLP-----EGAPQALDITQAP--KTPRANLQEQI 395
DB 334 LLLITGPPS-----DPAQAPALPGVLVSTEDVIDATERITIMTFINDLQMDM 384
QY 396 KAGKATYDSNCAACHQPDGKGVNAPPLANSDYLNADHARA 438
DB 385 QAVDVVIDHD-----QPEWIVITN-----DNSDMENFHVHDA 417

```

```

RESULT 9
US-10-115-563-14
; Sequence 14, Application US/10115563
; Publication No. US20030008307A1
; GENERAL INFORMATION:
; APPLICANT: Griffin, John H
; Greengard, Judith S
; TITLE OF INVENTION: METHODS FOR DIAGNOSING ACTIVATED PROTEIN
; C RESISTANCE ASSOCIATED WITH A FACTOR V GENETIC MUTATION
; AND COMPOSITIONS THEREOF
; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: The Scripps Research Institute, Office of

```

```

; Patent Counsel
; STREET: 10666 No. US20030008307A1ch Torrey Pines Road, TPC 8
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/115,563
; FILING DATE: 02-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/410,488
; FILING DATE: 24-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Filting, Thomas
; REGISTRATION NUMBER: 34,163
; REFERENCE/DOCKET NUMBER: 449.0
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-554-2937
; TELEFAX: 619-554-6312
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2224 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 14:
US-10-115-563-14

```

```

Query Match      4.9%; Score 128.5; DB 15; Length 2224;
Best Local Similarity 19.7%; Pred. No. 0.036;
Matches 82; Conservative 61; Mismatches 150; Indels 123; Gaps 20;
QY 63 DAIVTHAPEVPPVRDRHAKVIVKMET-----VEKVMKLADGVEYQFTFGQ----- 111
DB 358 EYIMDYAPVTPANMDKYSQHLDNFSNOIGHYKKVM-----TQYEDSFTKHTVPPM 413
QY 112 ----VPGMIRVREGDTIEVQFSN--HPSKMPHNVDFAATGPGGAESFTAPGHTS 164
DB 414 KEDGILGPIIRAQVBDTLKIVFKMASRPSYSLPHGVTSPYEDE--VNSSFTS-GRNN 469
QY 165 TFSFKALPG-LVYV-----HCAVAP-----VGMHIANGMGLIYVER 202
DB 470 TM-IRAVQGETYTYKMNILBEPDEPTENDAOCLTRPYSDVDIMRDIAGLIGLLICKS 528
QY 203 EGLPK-----VDKEYVYVQGF-YTKGKYGEGLOPF--DMBKAIREDAEVYVFNVS 253
DB 529 RSLDRGIRAADIEQAVFAVFDENKSWYLEDNINKFENDEYKRDPPK----- 580
QY 254 ALTGENALAKAVGETVRLFYNGGPNLTSFHVIGEIFD-----KVAF 296
DB 581 -----YESNIMSTINGV-----PESITTLGFCFDQDVQWMCVSTQNEILTIHF 626
QY 297 EGK--GENNNIQTLLIPAGGAITEFVVDVPGDYVLVDHAI FRAFNKALGILKVEE 353
DB 627 TGHSPFYGRHEDTLTLPFRGSESVT-----VTMD-----NVGTWMLTSMNS 669
QY 354 ENHEIYSHKQTDAAVLPBGAQALDITQAPKTPAPA-----NLOEQIKAGATYD 403
DB 670 PRSKLRLEKFDVKCIPDDDEDSYEIPEPESSTWATRKMDRLPEDEESDADYD 725

```

```

RESULT 10
US-10-172-712-31
; Sequence 31, Application US/10172712
; Publication No. US20030125232A1
; GENERAL INFORMATION:
; APPLICANT: GRIFFIN, JOHN H.

```

APPLICANT: GALE, ANDREW J.  
APPLICANT: GETZOFF, ELIZABETH D.  
APPLICANT: PELLEGUER, JEAN-LUC  
TITLE OF INVENTION: STABILIZED PROTEINS WITH ENGINEERED DISULFIDE BONDS  
FILE REFERENCE: 4198-4001US1  
CURRENT APPLICATION NUMBER: US/10/172,712  
CURRENT FILING DATE: 2002-09-30  
PRIOR APPLICATION NUMBER: 60/298,578  
PRIOR FILING DATE: 2001-06-14  
NUMBER OF SEQ ID NOS: 32  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 31  
LENGTH: 2224  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-172-712-31

Query Match 4.9%; Score 128.5; DB 15; Length 2224;  
Best Local Similarity 19.7%; Pred. No. 0.036;  
Matches 82; Conservative 61; Mismatches 150; Indels 123; Gaps 20;

QY 63 DAIVTHAEVPPYDRDHPAKVAVVVKMET-----VEKWRLADGVEYQFMTFGGQ----- 111  
DB 358 EVIWDVAPVPIRANMDKRTSGHLDNFSNQIGKHKVMY-----TQYDESFTHKTVNPNM 413  
QY 112 ----VPGQMRVREGDTIEVOFSN---HPDSKMPHNVDFAHATGPGGABASTAPGHTS 164  
DB 414 KEDGIIIGPIRAQVRDITLKIIVKMAGRPYSIYPHVGTFSPYEDE---VNSSFTS-GRNN 469  
QY 165 TFSFKALQF-LVYV-----HCAVAP-----VGMHANGYGLILVEPK 202  
DB 470 TM-IRAVQGETYTKWNILFDEPTENDACLTFRPYSDVDIMRDASGLIGLLICKS 528  
QY 203 EGI.PK-----VPEKEYVMQGF-YTKGYGEOGLQGF--DMEKALIEDAAYVFNQSVG 253  
DB 529 RSLDRRIGQAADIEQAVAVFPEENKSWLEDNINKFCENPDEVKRDKPF----- 580  
QY 254 ALTGEMALKAKVGETVRLFVNGGPNLTSSPHVIGELFD-----KVHF 296  
DB 581 -----YESNINSTINGV-----PESITTLGFCFDDTVQWFCVGTQNEILTIHF 626  
QY 297 EGCG--GENHNITOTLIPAGGAITEFKVDVPGDYLVVHAIFRANKALGLIKVEGE 353  
DB 627 TGHSEFYKQHEHDTLTLFPMRGESVT-----YTMV-----NVSTMWLTSMNS 669  
QY 354 ENHEIYSHKQTDVAVLPEGAPOAIDTQAPKTPA-----NLQEOIKAGKATYD 403  
DB 670 PRSKKRLKFRDVKCIFDDEDSYEIFEPPESTYMAIRKMHDRLEPEDESADYD 725

RESULT 11  
US-09-732-350-7  
Sequence 7, Application US/09732350  
Patent No. US20010031490A1  
GENERAL INFORMATION:  
APPLICANT: Svendsen, Allan  
APPLICANT: Xu, Feng  
TITLE OF INVENTION: LACCASE MUTANTS  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. US20010031490A10 No. US20010031490A1disk of No. US20010031490A1ch  
STREET: 405 Lexington Avenue  
CITY: New York  
STATE: NY  
COUNTRY: USA  
ZIP: 10174  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/732.350

FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/032,315  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Rozek, Carol  
REGISTRATION NUMBER: 36,993  
REFERENCE/DOCKET NUMBER: 5200,200-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-867-0123  
TELEFAX: 212-878-9655  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 572 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-732-350-7

Query Match 4.7%; Score 124; DB 9; Length 572;  
Best Local Similarity 21.4%; Pred. No. 0.011;  
Matches 79; Conservative 55; Mismatches 124; Indels 112; Gaps 20;

QY 86 VKMETVEKWRLA-----DGVEYQFMTFGGQVPGQMRVREGDTIEVOFSN---HPDSK 136  
DB 17 VLARTVEYNLKIISNGKIADGVERDATLVNGGYPGLIFANKGDTLKVKYQNKLTNPDMY 76  
QY 137 MPRNVDFHAA-----TGPGGABASTAPGHTSFSFK-ALQGLYVYHCANAPVGM 187  
DB 77 RTTSHIMHGLQHRNADDGPAFVITQCPV-VQASATYTMPGLDQGTGYWH---SHLS 132  
QY 188 HIRANGMYG-LILVEPEKGPVKV---DKEYVMQGFY-TKGG---YGEQGLQPFDMK 237  
DB 133 QYVDGRGLVLYDPDPRRLYDIDDEKTVLLIGDWYTSKALATGNTLQCPDS-- 190  
QY 238 AIREDAAYVFNQSVGALGTEN-----ALKAKVGETVRLFVNGGPNLTSSPHVIG 288  
DB 191 -----ATINGK-GRFPDNTPANPNTLYTLKVRGRKRYLRVINS--SALASF----- 235  
QY 289 ELFDKVFHFGCGENHNITOTLIPAGGAITEFKVDVPGDYLVVHAIFRANKALGLIL 348  
DB 236 ---RMSIQGHK-----MTVIADGVSTKPYVD----- 260  
QY 349 KYEGSENHIEYSHKQTDVAVLPEGAPOAIDTQAPKT---PAP-ANLQEOIKAGKATYDS 404  
DB 261 -----SPDILAGQRIDAV-----VEANOEPDTYWINAPLTVANKTAQALLIYED 305

RESULT 12  
US-09-869-877-7  
Sequence 7, Application US/09869877  
Publication No. US20020192792A1  
GENERAL INFORMATION:  
APPLICANT: Schneider, Palle  
APPLICANT: Danielson, Steffen  
TITLE OF INVENTION: Laccase Mutants  
FILE REFERENCE: 10179,204-US  
CURRENT APPLICATION NUMBER: US/09/869,877  
CURRENT FILING DATE: 2001-07-06  
NUMBER OF SEQ ID NOS: 10  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 7  
LENGTH: 572  
TYPE: PRT  
ORGANISM: Rhizoctonia solani  
US-09-869-877-7





```

OY 269 RRLFGNGSPNLTSSHVHGEIFDKXHPFGKSEKNNICOTLLPAGGAATKRVVPGD 358
Db 215 MRLISLSCPNN--QPSIDGHELTTEVBDQLTEPTVDRLOLFTGORISFVIDANOP-- 270
OY 359 YVLVDHAFRAF--NKGALGILKVGEGEHNHIYSHKOTDAYLLEPGAPOAIDTOEAPTPA 387
Db 271 ---VDNYWYRAQPNKGRNGLAGT-----FANGVNSAILRVAAGAAANDPTTSA--NPN 317
OY 388 PANLOEQIAGKATYDSCNCAHQDPQKGVNPAFFPLANSDYINADHARAASIVANGLSG 447
Db 318 PAQLNE-----ADHLALIDPAAPGIPPT--PGAAD---VNLRFQJGFS-----CG 356
OY 448 KITVNGNOYESVMPAI-----ALSDQOIAN 472
Db 357 RFTINGTAYES--PSVPTLLQITNGASQAN 384

```

```

RESULT 15
US-09-338-723A-2
; Sequence 2, Application US/09338723A
; Patent No. US20020019038A1
; GENERAL INFORMATION:
; APPLICANT: Huaming, Wang
; TITLE OF INVENTION: Phenol Oxidizing Enzymes
; FILE REFERENCE: GC561-2
; CURRENT APPLICATION NUMBER: US/09/338,723A
; CURRENT FILING DATE: 1999-06-23
; PRIOR APPLICATION NUMBER: 09/720,871
; PRIOR FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 594
; TYPE: PRT
; ORGANISM: Stachybotrys chartarum
US-09-338-723A-2

```

Query Match	4.4%	Score 114.5	DB 9	Length 594
Best Local Similarity	19.0%	Pred. NO. 0.092		
Matches 123	72	Mismatches 226	Indels 225	Gaps 28

```

Oy 6 I I K T L I C A L S L M J S G S N O A D K K A O K S S I V D A A A T --- A N D N A S O H Q E L V I 62
Db 2 L F K S Q L A A A S G L S G V G I P M D T S H P - I E A N D E V K T E V P A D S L L A A A G D D M E S P Y 60
Oy 63 D A I V T H A E V P P ----- P V - D R D --- H P A K V V K M E T E V K M E L A D G V E Y Q F W T F 108
Db 61 N L Y N A N L P I P P V K O P K M I I T P V Y G K D I M Y E I E I K P F O O R I Y P T L P A T L V G D --- 116
Oy 109 G G O V G C M I R A R E G G T I E V O P S N H D S K P H N V D H A --- A O P G G A A S F T A R H T S T F 166
Db 117 - G M S G P P F N P R G E T V R F T N - N A T E N S V H L G S P S R A P F D G M A E D V T F P G E Y K D Y 173
Oy 167 S F K A L O P G - L Y V Y H C A V A P V G M H I A N O M Y G L I L V --- E P K E G L P K V D E K Y V Y W Q G D F Y 220
Db 174 Y F P N Q S A R L L M Y H D H A F M K T A E N N Y F O Q A G Y I I N D A E A D A L G P --- 219
Oy 221 T K G K G E O G L O P F D M E K A I R E D A E V V F N G S V G A L T G E N A ----- 260
Db 220 -- S G Y G E ----- F D I P L I L - T A K Y N A D G L T R S T E G D O D L M G D V I H V N G O P M F L N V Q 270
Oy 261 ----- L K A K V G E T V R L F - V G N G S P A L T S F H Y I G E ----- 289
Db 271 P R K T R F R L N A A V S A M L I L Y V R T S S P V R I P F O V I A S D A G L L O P V O T S N L Y L A V A E R Y 330
Oy 290 -- I F D K V H F E G K G E N H N I O T T ----- L I P A G ----- 314
Db 331 E I I D F T F A G Q T L D L R N V A E T N D V G D E D E V A R T L E V M R F V S G S T E D N S Q V E T L D V 390
Oy 315 ----- G A L T E F R V D V P G D V L V D H A I F --- R A N N K A L G I L K Y E G E N ----- H 356
Db 391 P F P P H K E P R A D K H F F E R S N G H Y L L I N D G F A D V N E R V I L A K E L G I V E A W E L E N S G G M S H 450

```

QY 357 ELYSH-----KQT-----DAYLPEBAPQALDTGAAPTTPAPA-----N 390  
:::  
Db 451 PVAIHILVDFLIKRTGGRGQWMPYESAGLKDVMLGRETLLTTEAHYPQMTGAVMHCHN 510  
  
QY 391 LGEQIKAKRATYDSCAACHQ-----PDGGGVNAFPPLANSDYLNAD--HAR 436  
:::  
Db 511 LIHEDDMMAVF--VVTLMEEKGYIQEDFEDPMNPKNMAYF-----YNRNDFHAR 558  
  
QY 437 AASIVANGSGKITVNGQYESVMPAIALSDQAIANITYTLNSFG 482  
:::  
Db 559 AGNFSAESTIARVQ-----ELAQEPPYNRULDELIEUG 591

```

RESULT 16
US-10-080-210-2
; Sequence 2, Application US/10080210
; Publication No. US20020142423A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Huaming
; APPLICANT: Bodie, Elizabeth A.
; TITLE OF INVENTION: Phenol Oxidizing Enzymes
; FILE REFERENCE: GC561-3
; CURRENT APPLICATION NUMBER: US/10/080,210
; CURRENT FILING DATE: 2002-02-19
; PRIOR APPLICATION NUMBER: US 09/220,871
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: US 09/338,723
; PRIOR FILING DATE: 1999-06-23
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq For Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 594
; TYPE: PRT
; ORGANISM: Stachybotrys chartarum
; US-10-080-210-2

```

Query Match	4.4%	Score	114.5	DB	14	length	594
Best Local Similarity	19.0%	Pred. No.	0.092				
Matches	123	Conservative	72	Mismatches	226	Indels	225
						Gaps	28

QY	6	LIKTLLICALSAMLSCGCSNOADKAAOPKSTVDAAKTT---ANADNAAASGHOGLPEVI	62
Db	2	LFSKMWQLAASGLISGLVGLIPMDYGNRP-IEAVDPEKTEVFADSLAAAGDDDDWESPYY	60
QY	63	DAIVTAHAEVPP-----PV-DRD---HPKRVVVKQETVEKVMRLADGVEYQFWTF	108
Db	61	NLYRNALPIRPVQKPMIITNPATGDIWYEIEIRPFQORIIYPTLRPALTVGDD-----	116
QY	109	GGGVPGGMIVREBDTIEVGFPSNHPDSKMHNVDFHA--ATPGCGAASATPAGHSTF	166
Db	117	GMSPGSTEVPRTETVVRPFIN--NATVEISVHLHSSPSPAPDPGMAEDVTFPEEYKDY	173
QY	167	SFKALQPG--LYVYHCVAAPVGMHINANGYLILV---BEKEGLPKVDKEYVYVWQGDIFY	220
Db	174	YFPNYSARLLWHHDHAFMKTAENAYRFGQAGAYIINDEADALGP-----	219
QY	221	TKGKYEQGLQOPEDMEKATREDAEYVFNFSVCAITGEMA-----	260
Db	220	--SGYGE-----FDIPILIT--TAKYVYADDTLRTEEBDDDLGWDVIHVNGQWPFNLVQ	270
QY	261	-----LKAQGETVRLF--VNGGCGNLTSSPFVIGE-----	289
Db	271	PKRRFRFLNAAASRAWLLIYLVRTSSENVNAPIQVIASDAGLLQAPQTSNLYLVAVERV	330
QY	290	--IFDKVHEFGKGENHIQTT-----LIPIAG-----	314
Db	331	ELIITDFNFGQTLDLNNVAETNDVGDEDEYARTLEWRFVSSGTEEDNSQVPSLIRDV	390
QY	315	-----GAATFEKVDVDPGDYLVLDHAIF-----PAFNKGAGILIVVEGSE-----H	356
Db	391	PPRPHEGPRDKHFKPKERSNGHYIINDVGADVNERVLARPELGTVEWELENSSGGMWSH	450

QY 357 ELYSH-----KOT-----DAVYLBEPARQALDIOEAKTPAPA---N 390  
DB 451 PVHILVDEKIKRTGRCQVMPYESAGLKDYYWLGREGTTLTEAHYQWTAQYMHCHN 510  
QY 391 LOEOIKAGKATYDSNCACHQ-----PDGKGVNAPFPLANSDYLNAD-HAR 436  
DB 511 LIHENDMMAVF--NTVAMEEKGYLOEDFEDPMNPKRAVP-----YNRDFFHAR 558  
QY 437 AASIVANGLSGKITVNGQYESWPAIALSDOOIANVITYTINSFG 482  
DB 559 ACFNSASBITARVQ-----ELAEQEPYRRLDEILEDG 591  
RESULT 17  
US-10-080-233-2  
; Sequence 2, Application US/10080233  
; Publication No. US20020151450A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Huaming  
; TITLE OF INVENTION: No. US20020151450A1e1 Phenol Oxidizing Enzymes  
; FILE REFERENCE: GCS67  
; CURRENT APPLICATION NUMBER: US/10/080,233  
; CURRENT FILING DATE: 2002-02-19  
; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 594  
; TYPE: PRT  
; ORGANISM: Stachybotrys sp.  
US-10-080-233-2

Query Match 4.4%; Score 114.5; DB 14; Length 594;  
Best Local Similarity 19.0%; Pred. No. 0.092;  
Matches 123; Conservative 72; Mismatches 226; Indels 225; Gaps 28;  
QY 6 LIKTLICLSALMSGCSNOADKAAQPKSSTYDAAKT---ANADNASQEHQELPVI 62  
DB 2 LFKSQQLAASGLISGLVGLIPMDTSGHP-IEAVDPEVKTEVPADSLAAGDDWESPPI 60  
QY 63 DAIVTHAPEVPP-----PV-DRD---HPAKVVMGEMTEKVMRLADGVEYQFWTF 108  
DB 61 NLTYNLAALPFPVQKPKMILITNPVTKDLYEIEIKPQOKRIPTLRPATLVGD---- 116  
QY 109 GGOVPGOMIRVREGDTIEVOFSNHPDSKMPHNVDFHA--ATGPGGAEASFTAPGHTSTF 166  
DB 117 -GMSGPFTFNVPRGRETIVRFIN--NATVENSYHLHGSPSRAPFQDMADVTFFPGYKXY 173  
QY 167 SFKALQPG--LYVHCAYAPVGMHIANGMVGLIV---EPKEGLPKVDKEYYVWQGDYF 220  
DB 174 YEPNYQARLLMYHHAFFMKTAEENAVFGQAGAVIINDAEADALGLP----- 219  
QY 221 TKGSGEGSLOPFDMEKAIREDAEVYVFNQSGALGTENA----- 260  
DB 220 -SGYGE-----FDLPILV--TAKYNNADGTLRSTEGEDQIDMGVIVHNGQWPPLNQ 270  
QY 261 -----LKAKYGETVRLF--VGNGGPNLTSSFHVG----- 289  
DB 271 PRKYFRPLNAAVSRAAMLLYVTRTSSPNRIFPQVYASDAGLLQAPVQTSNLYLAERY 330  
QY 290 --IDPKVHEGSGENHNIQTT-----LIPAG----- 314  
DB 331 EIIIDFTFAGQTLRLRNVAETNDVGDDEYARTLEVMRFVSSGTVEDNSQVPSFLRDV 390  
QY 315 -----GALTEFVYDVGDVYLVNDHALF-----RAFKKALGLIKVGEEN-----H 356  
DB 391 PPPPKHEGADGHNFPERSNHLYLNDVGFADVNEVLAKPELGTVEVLELSSGSGWSH 450  
QY 357 ELYSH-----KOT-----DAVYLBEPARQALDIOEAKTPAPA---N 390  
DB 451 PVHILVDEKIKRTGRCQVMPYESAGLKDYYWLGREGTTLTEAHYQWTAQYMHCHN 510  
QY 391 LOEOIKAGKATYDSNCACHQ-----PDGKGVNAPFPLANSDYLNAD-HAR 436

DB 511 LIHENDMMAVF--NTVAMEEKGYLOEDFEDPMNPKRAVP-----YNRDFFHAR 558  
QY 437 AASIVANGLSGKITVNGQYESWPAIALSDOOIANVITYTINSFG 482  
DB 559 ACFNSASBITARVQ-----ELAEQEPYRRLDEILEDG 591

RESULT 18  
US-09-995-749A-13  
; Sequence 13, Application US/09995749A  
; Patent No. US20020155568A1  
; GENERAL INFORMATION:  
; APPLICANT: VAN GEEEL-SCHUTTEN, GERRITDINA HENDRIKA  
; APPLICANT: DIJKHUIZEN, LOBBERT  
; APPLICANT: RAHAOUI, HAKIM  
; APPLICANT: LEER, ROBERT-JAN  
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASES  
; FILE REFERENCE: BO43388-CIP  
; CURRENT APPLICATION NUMBER: US/09/995,749A  
; PRIOR FILING DATE: 2001-11-29  
; PRIOR APPLICATION NUMBER: 09/604,957  
; PRIOR FILING DATE: 2000-06-28  
; PRIOR APPLICATION NUMBER: EPO 00201871.1  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 13  
; LENGTH: 535  
; TYPE: PRT  
; ORGANISM: Lactobacillus reuteri  
US-09-995-749A-13

Query Match 4.3%; Score 113.5; DB 10; Length 535;  
Best Local Similarity 21.8%; Pred. No. 0.097;  
Matches 73; Conservative 43; Mismatches 116; Indels 103; Gaps 16;  
QY 165 TFSFKALQPLGLVYV---HCAVAPVGMHIANGMVGLILVEPKGLPKVDKEYYVWQGDYF 220  
DB 175 TTFPEDEQKIDAYIADONSTVKKYNLNIPASVALILTN-KDTIPRV---GDLY 227  
QY 221 TK-GKYGEGLOPFD-MEKAIREDAEVYVFNQSV-----GALTGN 259  
DB 228 TDGQGYMEHQRTFYTYTLTNLKSRYKVVAGQSMQMTMSGVGNMNLTSRYGKGAMTARD 287  
QY 260 A-----LKAKYGETVRLFVGNGGPN-----LTSSFHVGIFPK 293  
DB 288 TGTDETRQIGCVVSNTPNLKLVNDKVLHMGAAHAKQOYRAVAVLTTTGDVINTSDQ 347  
QY 294 VHEGKGENHNIQTTLIPAGGAITTEKVDVPGDVYLVNDHALFRAFNKALGLIKVEGE 353  
DB 348 -----GAPVA--MTDENGDLYSSNHL--VNGKEADDTAVQGY 382  
QY 354 ENHEIYSHKQTDVYLBEPARQALDIOEAKTPAPANLOEOIKAGKATYDSNCACHQPD 413  
DB 383 ANPDVSGYV---AVWVPVQASDNDQARTAPST-----EKSGNSAVRTNAF---D 427  
QY 414 GKGVENAF-----PLANSDYLNADHARAASIVAN 443  
DB 428 SNVITFAFSNFTYPTKESERANVRIQWADFFAS 462  
RESULT 19  
US-09-995-749A-2  
; Sequence 2, Application US/09995749A  
; Patent No. US20020155568A1  
; GENERAL INFORMATION:  
; APPLICANT: VAN GEEEL-SCHUTTEN, GERRITDINA HENDRIKA  
; APPLICANT: DIJKHUIZEN, LOBBERT  
; APPLICANT: RAHAOUI, HAKIM  
; APPLICANT: LEER, ROBERT-JAN  
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASES  
; FILE REFERENCE: BO43388-CIP  
; CURRENT APPLICATION NUMBER: US/09/995,749A

CURRENT FILING DATE: 2001-11-29  
PRIOR APPLICATION NUMBER: 09/604,957  
PRIOR FILING DATE: 2000-06-28  
PRIOR APPLICATION NUMBER: EPO 00201871.1  
PRIOR FILING DATE: 2000-05-25  
NUMBER OF SEQ ID NOS: 19  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 2  
LENGTH: 1781  
TYPE: PRF  
ORGANISM: Lactobacillus reuteri  
US-09-995-749A-2

Query Match 4.3%; Score 113.5; DB 10; Length 1781;  
Best Local Similarity 21.8%; Pred. No. 0.63;  
Matches 73; Conservative 43; Mismatches 116; Indels 103; Gaps 16;

QY 165 TFSFKALQPLGVY---HCAVAPVGMHANGYLIVPEKGLPRKDEYVVMQDFY 220  
DB 1154 TTFEEDQKIDAVIDQNSTVKKYNLNPASAYAILLTN-KDTIPRV---YY---GDLV 1206  
QY 221 TK-GKXGEQLOQFDP-MEKAIREDAYVNVGSV-----GALTGEN 259  
DB 1207 TDGQYMEHQTRYDVLTLNLKSRKXVAGQSWQTSVGGNNNLTSTVYGGKAMATPD 1266  
QY 260 A-----LKAKEGTVLFPVNGGPN-----LTSFHYIGEIFDK 293  
DB 1267 TGTDETRTQIGVNVSTPLKGVNDKVLHGAHKNQYRAAVLTITDGIYNTSDQ 1326  
QY 294 VHEEGKGENHNIQTLLIPAGAAITFEKVDVPGDYLVDHAI FRAENKALGILKEGE 353  
DB 1327 -----GAPVA--MTDENGDTLSSHNL--VYNGKEADTAVQGY 1361  
QY 354 ENHEIYSHKOTDVLPEGAPOALDTQEAARKTAPAPALQOIQIAGKATYSNCAQHPD 413  
DB 1362 ANPDVSGYL---AVWPFVGSNDQDARTAST-----EKSGNSAYRTNAF---D 1406  
QY 414 GKGVNPAF-----PPLANSDYLANDHARAASIVAN 443  
DB 1407 SNVIFFAFSNPFVYTPTESERAVRRIAQNADFPAS 1441

RESULT 20  
US-10-095-718-4  
Sequence 4, Application US/10095718  
Publication No. US20020131956A1  
GENERAL INFORMATION:  
APPLICANT: Walsh, Christopher  
APPLICANT: Chao, Hengjun  
APPLICANT: Bursstein, Haim  
APPLICANT: Lynch, Carmel  
APPLICANT: Stepan, Tony  
APPLICANT: Munson, Keith  
TITLE OF INVENTION: Adeno-Associated Virus Vectors Encoding Factor VIII and  
FILE REFERENCE: 35052/204375  
CURRENT APPLICATION NUMBER: US/10/095, 718  
PRIOR FILING DATE: 2002-03-12  
PRIOR APPLICATION NUMBER: 09/689,430  
PRIOR FILING DATE: 2001-08-22  
PRIOR APPLICATION NUMBER: 60/158,780  
PRIOR FILING DATE: 1999-10-12  
NUMBER OF SEQ ID NOS: 5  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 4  
LENGTH: 1431  
TYPE: PRF  
ORGANISM: canine B-domain deleted factor VIII  
US-10-095-718-4

Query Match 4.3%; Score 113; DB 14; Length 1431;  
Best Local Similarity 20.3%; Pred. No. 0.5;  
Matches 82; Conservative 66; Mismatches 154; Indels 102; Gaps 22;

QY 68 HAPVEPPVDRDH-----PAKVVMKMEV-----EKWRLADGVEVFWTFGGQVP 113  
DB 408 YABSGPTPDRSHKULYLANGPQRIQKTKYKRPAAVYTBETRTREAIQYESGILGPLLY 467  
QY 114 GQMIKVRREGDTIEVOFSN--HPDSKMPHNVDF---HAATPGGGAESAF--TAGHTS 164  
DB 468 GEV-----GDTLLIIFKNQASRPVNIYPHGINVVPPLHTGRLPKGVKHKDMPILFG--E 520  
QY 165 TFSFK--ALQPG-----LVTHCAVAPVGMHANGYLIVPEKGLPK----- 207  
DB 521 IFKRYKWTVEEDGPTKSDPRCLTRYSSPINERDLASGLIGPLLCYSESVDORGNQM 580  
QY 208 VPKKEYVMQGF-YTKGKXGEQLOPF--DMEKAIREDAYVFN--GSVGLTGEN-AL 261  
DB 581 SDRKAVILSVSDENRSMWLTENMORFLPNADVQPHDEPQLSNMHSINGVFPNLOJ 640  
QY 262 KAKVGETVLFPVNGGPNLTSFHYIGELFDKVEEGKGENHNI---QTLIPAGAAI 318  
DB 641 SVCLHEVAVAYILSVQAQTD-----FLSVFSGYTFKKHMYVEDTLTLPFGGETV 691  
QY 319 TEKVDVPGDYLVDH-AIFRAPKAGLILKEGSEHHEIYSHKOTDVLPEGAPOAI 377  
DB 692 F-MSMENPGLWVLGCHNSDFR--NRGMTALIKVSS-----CNRNIDDY-----E 733  
QY 378 DTQEAKEPTA-----PANTLOEIKAKATYD 403  
DB 734 DTYEDIPTPLANNVNIKPRFSQNSRHPSTKEKOLKMRBDP 777

RESULT 21  
US-10-174-693-290  
Sequence 290, Application US/10174693  
Publication No. US2003013173A1  
GENERAL INFORMATION:  
APPLICANT: Bloksberg, Leonard N.  
APPLICANT: Havukkala, Ilkka  
TITLE OF INVENTION: Modification and Methods for the  
FILE REFERENCE: 11000.100365  
CURRENT APPLICATION NUMBER: US/10/174, 693  
PRIOR FILING DATE: 2002-06-18  
PRIOR APPLICATION NUMBER: US 08/975,316  
PRIOR FILING DATE: 1997-11-21  
PRIOR APPLICATION NUMBER: US 08/713,000  
PRIOR FILING DATE: 1996-09-11  
PRIOR APPLICATION NUMBER: US 09/169,789  
PRIOR FILING DATE: 1998-10-09  
PRIOR APPLICATION NUMBER: US 09/615,192  
PRIOR FILING DATE: 2000-07-12  
NUMBER OF SEQ ID NOS: 407  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 290  
LENGTH: 440  
TYPE: PRF  
ORGANISM: Eucalyptus grandis  
US-10-174-693-290

Query Match 4.2%; Score 110.5; DB 16; Length 440;  
Best Local Similarity 19.7%; Pred. No. 0.14;  
Matches 86; Conservative 51; Mismatches 136; Indels 163; Gaps 20;

QY 111 QYPGMIRVREGDTIEVQFSNHPDSKMPHNVDF-----ANTPGGGAESAFTAPGHS 164  
DB 75 QYGPPLVAREGDNILVKNVH---VAANVTIHHGVROLRTGMADGAVYTQCTIQTN 130  
QY 165 ---TFSFKAL--OPGLVYVHCAVAPVGMHANGYLIVPEK--EGLP--KYDKEYVWQ 216  
DB 131 QSYTNFTLTGQRTGLTMAHVS---WRRSIHGPITLIPKRNESYPRKESKVEPIIF 186  
QY 217 GPFYTKGKXGEQLOPFDMEKAIRED-----AEYVFNQSVGALTGENA-----LK 262  
DB 187 GEFMFN-----VPEAVIAQALQSGGPNVSDATYINGLPGPLVNCSSKDTFKLK 235

QY 263 AKYGEFVRLFGVNGGPNLTSSPHV-----IGEIFDK-----VFEGKGENHNIO 307  
 Db 236 VPRGKYLLRLINAAINDELFFESIANHATVVEVDVATYKPFSGACILHLPQGTMMVLK 295  
 QY 308 TLLIPAGGAITFEKYVDVGVYLVDAIFRAF-----NKALGILKEESEN----- 355  
 Db 296 T-----KTDFFPNSTFLM--AAMPYFTGMDSTVAGILEYEHKSSNYP 340  
 QY 356 -HEIYSHKOT-----DAVYLEGAPQALD-----TOBAPKTPAPA 389  
 Db 341 LKKLPQYKTLPPMNSTGVPAKFTGQLRSLASAKFPAANPQKDKRFFFTVGLGTSPCPK 400  
 QY 390 NLQEOIKAKATYDSNCAACHOPDGKGVNAPFPPLANSDYLNADHARAASIVANGLSGKI 449  
 Db 401 N-----TTCQGPNGT-----KF 412  
 QY 450 TVNGNOYESVMPAL 465  
 Db 413 AASVNNISFVLPSVAL 428

RESULT 22

US-09-118-276-2  
 ; Sequence 2, Application US/09118276  
 ; Patent No. US20010011381A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: BABIYCHUK, ELENA;  
 ; APPLICANT: KUSHNIR, SERGEI;  
 ; APPLICANT: DE BLOCK, MARC;  
 ; APPLICANT: INZE, DIRK  
 ; TITLE OF INVENTION: METHODS AND MEANS TO MODULATE PROGRAMMED  
 ; NUMBER OF SEQUENCES: 21  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: SIXBEY, FRIEDMAN, LEBDOM, & FERGUSON  
 ; STREET: 8180 GREENSBORO DRIVE, SUITE 800  
 ; CITY: MCLEAN,  
 ; STATE: VIRGINIA  
 ; COUNTRY: USA  
 ; ZIP: 22102  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: 3-1/2" DISKETTE  
 ; COMPUTER: IBM-COMPATIBLE  
 ; OPERATING SYSTEM: MS-DOS  
 ; SOFTWARE: ASCII  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/118,276  
 ; FILING DATE: 17-JUL-1998  
 ; PRIOR APPLICATION DATA: NONE  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: SCHULMAN, ROBERT M.; SALEN, KENNETH H.  
 ; REGISTRATION NUMBER: 31,196; 43,077  
 ; REFERENCE/DOCKET NUMBER: 6201-0003  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (703) 790-9110  
 ; TELEFAX: (703) 883-0370  
 ; INFORMATION FOR SEQ. ID NO. 2:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 969 RESIDUES  
 ; TYPE: AMINO ACID  
 ; STRANDEDNESS: SINGLE  
 ; TOPOLOGY: LINEAR  
 ; US-09-118-276-2

Query Match 4.2%; Score 109; DB 9; Length 969;  
 Best Local Similarity 18.8%; Pred. No. 0.64;  
 Matches 119; Conservative 72; Mismatches 188; Indels 254; Gaps 25;  
 QY 24 SNQADAAQPKSTVVA-AAKTANAMMASQ---EHQGLPVIDALVTATAPPEPPVDND 79  
 Db 175 NNEQNGSKRKXSENDIDSTYKSRLESTSEGTVRNKGQL-----VDPKGSNT 222

QY 80 HPAKVTVKM-ETVEKVMRLADGEVEYQFTFGQVQPCQMRVREBGTIEVQFSNHPDSKMP 138  
 Db 223 SSADIQLKXKEOSDITLMLKDKGLK-----THVSAEIRDLMEANGQDTSGP 268  
 QY 139 --HNVDENHATGPGGAASFTAPGHTSTFSFKALOPGLVYVHCAPVGMHIANMGYL 196  
 Db 269 ERHLID-RCADG-----MIFGALGP-----CPVCANGMYNNQYQC 304  
 QY 197 -----ILVEPEGLPKVDEYVYVQGFYTKGKGEOGLQPFDEKA- 238  
 Db 305 SGVSEMSKCTYSATEPVAVKKKMQIPIHGTNDYLMKMKSKQVKKPERVLPPMSPEKSG 364  
 QY 239 ----- 238  
 Db 365 SKATQRTSLSSGLDKLRFVVGQSKKANEMIEKTLKAGANFYARVVKDIDCLIAQGE 424  
 QY 239 -IREDAEY-----VFNQSVGALTG-----ENALAKYGETVRLVNGN 276  
 Db 425 LQNEAEVKKARLKIPIVREGYIGCVKKNKMLPDLKYLENALESKSGSTVTVKVGK- 483  
 QY 277 GPNLTSSFHVGIEIFDKVH-FEGKGGENHIOITLIPAGGAATFEKVDVPGDYVLV--- 332  
 Db 484 ----RSAYHSSGLQDTNAILLEDGS-----IYNATLMSDLALGVNSYVQIIE 530  
 QY 333 --DHALFRAFKN-GALGILKEEENHEIYSHKQTDV----- 367  
 Db 531 ODDGSECVYFRKMGVRSKIGQKLEEM---SKTEALIEFKRLFLEKTNSEWEACEKT 587  
 QY 368 -----YLPEGAPOALDTPQAPKTPAPANQ-----EQIRAGKATYDS 404  
 Db 588 NRRKOPGRFPYPLDVVDGVKAPKARKDISEMKSLAPQLLELMKMLFNVTYTRAAWMEFEI 647  
 QY 405 NCAACHQPDGK---GVNAPFPPLANSDYLNADHARAASIVANGSGKITVNGNOYESV 460  
 Db 648 NMS--EMPLGKSKENIEKGFALTEIONLQDLTDQALAVESL---IVAASNREFTLI 702  
 QY 461 PAI-----ALSDQOJANYI 474  
 Db 703 PSIHPIHIEDDEDLMIKAKMLELQDIEIASKI 735

RESULT 23

US-09-118-276-11  
 ; Sequence 11, Application US/09118276  
 ; Patent No. US20010011381A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: BABIYCHUK, ELENA;  
 ; APPLICANT: KUSHNIR, SERGEI;  
 ; APPLICANT: DE BLOCK, MARC;  
 ; APPLICANT: INZE, DIRK  
 ; TITLE OF INVENTION: METHODS AND MEANS TO MODULATE PROGRAMMED  
 ; NUMBER OF SEQUENCES: 21  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: SIXBEY, FRIEDMAN, LEBDOM, & FERGUSON  
 ; STREET: 8180 GREENSBORO DRIVE, SUITE 800  
 ; CITY: MCLEAN,  
 ; STATE: VIRGINIA  
 ; COUNTRY: USA  
 ; ZIP: 22102  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: 3-1/2" DISKETTE  
 ; COMPUTER: IBM-COMPATIBLE  
 ; OPERATING SYSTEM: MS-DOS  
 ; SOFTWARE: ASCII  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/118,276  
 ; FILING DATE: 17-JUL-1998  
 ; PRIOR APPLICATION DATA: NONE  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: SCHULMAN, ROBERT M.; SALEN, KENNETH H.  
 ; REGISTRATION NUMBER: 31,196; 43,077  
 ; REFERENCE/DOCKET NUMBER: 6201-0003



```

; LENGTH: 2233
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-769-787-2

```

```

Query Match      4.1%; Score 108.5; DB 11; Length 2233;
Best Local Similarity 20.3%; Pred. No. 2.6;
Matches 110; Conservative 63; Mismatches 179; Indels 189; Gaps 30;

```

```

QY 58 ELPLVDIAVTHAPEVP-----PPVDRDHPAKVVK-----87
DB 994 EYPKVQTIIGAEPEPTTVPFYSDGSRAREPVTSSVDVSKPGIVTVKGMADGREVEAR 1053
QY 88 -----METVEKVMR-LADG--VEYQF--WTFGGQ-----VPG 114
DB 1054 VEVIKSELPLVKRIAPNTDLSVDKSVSYLLIDGSVEEYVDKMEIAEBDKAKLAIIFG 1113
QY 115 QMIRVR---EGD---TIEVQFSNHPDSKMPHNVDFHATGPGGGAELSFTA--PGHIST 165
DB 1114 SRIQATGYLEGPPIHATLVVEEGNPAPAPV-----TVTVGGEAVTGLTSQKPMQYRT 1166
QY 166 FSKFLOPGL--YVYHCAVAPVGMHIANGMYGLILVEPKEG-----LPKV--208
DB 1167 LAYGAKLPEVITASAKNAATVTVLQASANANGMASIFIQPKDGPLOTVYAIQFLBAPKIAH 1226
QY 209 -----DKEYVVMQGD--FYTKGKXGE-----QGLQPPDM 235
DB 1227 LSLQVEKADSLKEDQTVKLSVRAHYQDGTQAVLPADKVTFTSGE--GEVAIRKGMLELHK 1285
QY 236 EKAIREDAEYVFNFGSVGALITGENALKAKVGETVRLFVNGGNPNLTSSPHVIGEIFDKVH 295
DB 1286 PGAVTLNMEYEGAKQV--ELTIQANTEKKIAQSIR-----PVNVVTTDHOQPSLPATVT 1338
QY 296 FEGGKG--ENHIQITLLPAGGAATTEPKVDVPGDVVLVDHAIIFRAFNKGAILKVEGE 353
DB 1339 VEYDGFPTKHTVTQALP-----KEKLD-----SYQFFB--VLG--KVBCI 1376
QY 354 ENHEIYSHKQTDVAVLPFGAPQAIPTQEAAPTAPANLQEQIKAGKATYDSCAACHPD 413
DB 1377 D-LEARAKVSVGIVSEVSTTPIAEAPQLP-----ESVR-----TYSN--GHVS 1422
QY 414 GKGVNAPPLANSOYLADHARAASIVANGLSKITVNGNOYESVMA--IALSPOQIAN 472
DB 1423 AKVADAIRP-----EQYAKGVFTVNG-----RLEGTQLTKLHVRSAGQEQGAN 1469
QY 473 V 473
DB 1470 I 1470

```

```

RESULT 26
US-09-765-272-84
; Sequence 84, Application US/09765272
; Patent No. US20020061545A1
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigen and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESSES:
; ADDRESS: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/765,272
; FILING DATE: 22-Jan-2001

```

```

; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/361,083
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 84:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 722 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 84:
US-09-765-272-84

```

```

Query Match      4.1%; Score 108; DB 9; Length 722;
Best Local Similarity 19.8%; Pred. No. 0.5; DB 176; Indels 180; Gaps 27;
Matches 104; Conservative 65; Mismatches 176; Indels 180; Gaps 27;

```

```

QY 58 ELPLVDIAVTHAPEVP-----PPVDRDHPAKVVK-----87
DB 263 EYPKVQTIIGAEPEPTTVPFYSDGSRAREPVTSSVDVSKPGIVTVKGMADGREVEAR 322
QY 88 -----METVEKVMR-LADG--VEYQF--WTFGGQ-----VPG 114
DB 323 VEVIKSELPLVKRIAPNTDLSVDKSVSYLLIDGSVEEYVDKMEIAEBDKAKLAIIFG 382
QY 115 QMIRVR---EGD---TIEVQFSNHPDSKMPHNVDFHATGPGGGAELSFTA--PGHIST 165
DB 383 SRIQATGYLEGPPIHATLVVEEGNPAPAPV-----TVTVGGEAVTGLTSQKPMQYRT 435
QY 436 LAYGAKLPEVITASAKNAATVTVLQASANANGMASIFIQPKDGP-----478
DB 478 LAYGAKLPEVITASAKNAATVTVLQASANANGMASIFIQPKDGP-----478
QY 479 -----LQTVAIQ--FLBAPKIAHL-----SLQVEKADSLKEDQTVKLSV-----516
DB 516 -----LQTVAIQ--FLBAPKIAHL-----SLQVEKADSLKEDQTVKLSV-----516
QY 517 -----RAHYQD-----TQAVLPADKVT--FSTSGE--VAIRKGMLELHKPG 556
DB 556 -----RAHYQD-----TQAVLPADKVT--FSTSGE--VAIRKGMLELHKPG 556
QY 557 AVTLNMEYEGAKQVDELTIQANTEKKIAQSIRPVNVVTTDHOQPSLPATVT-----609
DB 609 AVTLNMEYEGAKQVDELTIQANTEKKIAQSIRPVNVVTTDHOQPSLPATVT-----609
QY 610 --YDGFPTKHTVTQALPKEDLSDYQTFEVLGKVEGIDLE-ARA-----KVSVE 656
DB 656 --YDGFPTKHTVTQALPKEDLSDYQTFEVLGKVEGIDLE-ARA-----KVSVE 656
QY 657 GIVSVEVSVTTPIAEAPQLPESVRTYSN-----GHVSSAKVA 695
DB 695 GIVSVEVSVTTPIAEAPQLPESVRTYSN-----GHVSSAKVA 695

```

```

RESULT 27
US-10-238-075-877
; Sequence 877, Application US/10238075
; Publication No. US20030148324A1
; GENERAL INFORMATION:
; APPLICANT: I.N.S.E.R.M.
; TITLE OF INVENTION: Polynucleotides which are of nature B2/D+ A- and which are isolat
; FILE REFERENCE: BLANDINE
; CURRENT APPLICATION NUMBER: US/10/238,075
; FILING DATE: 2002-09-10

```

PRIOR APPLICATION NUMBER: 0003145  
 PRIOR FILING DATE: 2000-03-10  
 NUMBER OF SEQ ID NOS: 1576  
 SOFTWARE: PatentIn version 3.1  
 SEQ ID NO 877  
 LENGTH: 1684  
 TYPE: PRT  
 ORGANISM: Escherichia coli  
 US-10-238-075-877

Query Match 4.1%; Score 108; DB 12; Length 1684;  
 Best Local Similarity 20.3%; Pred. No. 1.9; Index 224; Gaps 33;  
 Matches 127; Conservative 76; Mismatches 199; Indels 224; Gaps 33;

5 TLTKTLTLCALSLMT-----SGCSNQ--ADKAQPKSSTV 38  
 63 TLNSALTYAGNNLKFTBRNLHQHGDILADNSLWQKSSGANSSEIIIRSGNIETMG 122  
 39 DAAAKTA---NADMAASQEHQGLPVIDAVTHAPEVPPVDRDHPAKVVMKETEYKVM 95  
 123 DITMNTAHLNLSWDALISASH-----EVIKSSRGVISPVPENRMWGVVR----- 167  
 96 RLADGVEY--QFTPEGQVPGQ--MIRVEDDTIEVQSNFSDSKMHNVDFAATGPGG 152  
 168 --HDGVEYLAVMGKATVPEDEYRITGDTVTVAATGHA--ARISGADMHIRARLDN 224  
 153 ABAFTAPGHTSTFSFKAL-----OPGL-----YVHCVAVPMHIANGMVGLIVEPK 203  
 225 -EASFILAGSMTLSGDTLNLNQMGQSTTCKEYVWLA-----SD 263  
 204 GLPKVDKEYVNVQGDFFYTKGY-----GEOGLOPFMEKAIREDAEYVNGSV 252  
 264 SLPKA-----WTFEPWYVVRQVSPDATEAGTSPVGQYRAVISAAGDV--SASF 311  
 253 GALTGNALKAKYGEVRLVFNCG---PNLTS-----SFVIGETIPKVFEGKG- 301  
 312 ATDTGNTTVPRAAG-----GSGNTITVPSLNSLTPVTSQGSSEALNL--ESGTGI 361  
 302 ---EHNHQTTL--IPAG---GAATFEP-----KVDPGVYVL----- 331  
 362 TGPVWDALPDTLKDPGALSILYGAVSSTPLPSGNNGVFPSTDDSPILIVNPKLDG 421  
 332 ---VDHAIFRAFNKGAIGILKVEGEENHEIYSHKQTDVYLPE---GAPQALDTGEAPK 384  
 422 LKGVDSLF---AGLYDLRLMQPGE---APRETPVYIDKQFSGSYIDRLGLPK 472  
 385 TEPAPALQGIKAKGATYDSNCAACHQPDQKGPVNAFPPLANDIYNAD----- 433  
 473 E-----KYRFLGDAAFDT-----RYVSNVILNQGTSGRYINGTGSDDLTKMKYLM 516  
 434 HARAASIVANGLSGKITVNGNQ-----YESV-----MPAIALSDQOI----- 470  
 517 DSAALAAQKALGLTFGSLTAGOVAQLTRSLMWESTTINGQYMWKLYISPDITLHNG 576  
 471 ---ANVTYTLNLSFGNKGGLSADD 492  
 577 SVISGNVQVLAGNITNSGGSINAON 602

RESULT 28  
 US-09-738-363-6  
 Sequence 6, Application US/09738363  
 Patent No. US20010010932A1  
 GENERAL INFORMATION:

APPLICANT: Schmeff, Harry E.  
 Schmeff, George E.  
 Payne, Jewel M.  
 Narva, Kenneth E.  
 Foncerada, Luis

TITLE OF INVENTION: Nematocidal Proteins  
 NUMBER OF SEQUENCES: 40  
 CORRESPONDENCE ADDRESS:  
 ADDRESS: Jay M. Sanders

STREET: 2421 N.W. 41st Street, Suite A-1  
 CITY: Gainesville  
 STATE: FL  
 COUNTRY: USA  
 ZIP: 32606  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA: US/09/738,363  
 APPLICATION NUMBER: US/09/738,363  
 FILING DATE: 15-Dec-2000  
 CLASSIFICATION: <Unknown>  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 09/076,137  
 FILING DATE: 12-MAY-1998  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Sanders, Jay  
 REGISTRATION NUMBER: 39,355  
 REFERENCE/DOCKET NUMBER: MA-20CCCD3  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 352-375-8100  
 TELEFAX: 352-372-5800  
 INFORMATION FOR SEQ ID NO: 6 (PS33F2):  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1257 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 HYPOTHETICAL: YES  
 ANTI-SENSE: NO  
 ORIGINAL SOURCE:  
 ORGANISM: Bacillus thuringiensis  
 INDIVIDUAL ISOLATE: PS33F2  
 IMMEDIATE SOURCE:  
 CLONE: E. coli NM522 (pMYC 2316) B-18785  
 FEATURE:  
 NAME/KEY: Protein  
 LOCATION: 1..1257  
 SEQUENCE DESCRIPTION: SEQ ID NO: 6:  
 US-09-738-363-6  
 Query Match 4.1%; Score 107.5; DB 9; Length 1257;  
 Best Local Similarity 19.2%; Pred. No. 1.3;  
 Matches 118; Conservative 69; Mismatches 200; Indels 227; Gaps 32;

25 NQADKAAQPKSSTVDAAKTANADNAASQEHQGLPVIDAVTHAPEVPPVDRDHPAKV 84  
 245 NGLDKF---KSLDVSYNKAKVYIKQMT-----MVLDLVA---LWTFEDPDHYQKE 290  
 85 VVKMETVEKVMADGVEYQ-----FWTFGGQVPGQMLRVNEDDTIEVQFSNHPDS--- 135  
 291 V---EIEPTRTISPI-YQVPVKMQNTSSSIVSDLFHY-QGDVLFKFSRTDNDGL 344  
 136 -----KMP-----HNVDFAATGPGGAAASFTAP----- 160  
 345 AKLFTGIRNTPFKSPHTHETVHVDPSYNTQSSGNSIRGSSNPIPIIDNNPIISTCIRNSF 404  
 161 -----GHTSTFSFKALQPGLYVYHCAVAPVG-----MHIANGY----- 194  
 405 YKALGSSVLVNFK---DGTGYAFQAPGTGAMDHSTFESDQAPRGHKLNTYTSFGPT 461  
 195 -----GLIVEPKGLPVRVDKEYVYMQG--DFTYTKGKGEGGLOPFME 236  
 462 LRDFINVTLLSTPINEISTEKIKGF-AEKGYLKNQGIKMYGKPEY-INGAQGVNHE 519  
 237 KA-----IREDAEYV-----FNGSVGLVTGENA-----LKAIVGET 268  
 520 NQOTLIEFHASKIAQYIRIRYASTQGTGKGFRLDNDGLQTLNPTSANGVYTNIGEN 579  
 269 VRLFVNGGPNLTSSPHVI-----GEIPDKVHFEGGKGHNHNIQTLLIPAGGAATTE 320



Db 580 YDLYT-IGSYTTEGNTTQIQHNDKNGWLDRIEF-----VPSDSIQ--- 621  
Qy 321 FKVDVGDVYLVDAIFRAFNKCALGLKVEGENEIIYSHKQTDVAVLPEGAPQADITQ 380  
Db 622 ---DSQDDEPPEVHESTIIFDKSSPTI---WSNHSYSHIHLESTYSQSS----- 667  
Qy 381 EAPKTPAPANLOEQIKAGATYDSNCACHQDPDGKVPNAPPL-ANSDYLNADHARAAS 439  
Db 668 -----YFHNLLINL-----FHPPTD-----PRNHTIHVNNQGMVNDYKDS- 703  
Qy 440 IVANGI-----SGKIT-----VNGQYESVMPALALSDQOIANVITYTLNS 480  
Db 704 -VADGINTFKKITATIPSDAMVSGTITSMHLENDNNFKITPKRELIN-ELENTITQVNAL 761  
Qy 481 FGNKGQSLADVA 494  
Db 762 FASSADOTLASNV 775  
  
RESULT 29  
US-10-043-487-383  
; Sequence 383, Application US/10043487  
; Publication No. US20030055220A1  
; GENERAL INFORMATION:  
; APPLICANT: Pierre, LEGRAIN  
; APPLICANT: HYBRIGENICS  
; TITLE OF INVENTION: Protein-protein interactions between Shigella Flexneri polypeptides  
; FILE REFERENCE: B4778A  
; CURRENT APPLICATION NUMBER: US/10/043,487  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/261,130  
; PRIOR FILING DATE: 2001-01-12  
; NUMBER OF SEQ ID NOS: 561  
; SOFTWARE: Patentin version 3.1  
; SEQ ID NO 383  
; LENGTH: 2609  
; TYPE: PRT  
; ORGANISM: Shigella Flexneri  
US-10-043-487-383  
  
Query Match 4.1%; Score 107.5; DB 15; Length 2609;  
Best Local Similarity 20.8%; Pred. No. 4.1;  
Matches 115; Conservative 71; Mismatches 217; Indels 149; Gaps 30;  
  
Qy 4 PRLITTLICALSALMLSCGNOADKAAQPKSSTVDAAKTANADNAAOEHQGLPVID 63  
Db 2113 PSVANVSHCHDLS-LKIPISIQ-DWTAQVTS---PSGKTHAEIIVEGENHTYCIHFVP 2166  
Qy 64 AIV-----THAPEVP-----PRVDRDHAKVNVKMEVEKVMFLADGV-EYGF 105  
Db 2167 AEMGHHTVSVKKGQVHPGSPPOFTVGLPGBEGANHVRAAGPGLF---PAEAGVPAPFESI 2223  
Qy 106 WTFGGVPGQMITREBGTIEVOFSNHPDS-----KMPHNVDFHAATGPGGAEASF 157  
Db 2224 WTRBAGAGLAIIVAGSPKAEISFEEDRKDSCGVAVVQEPDGYEVSVFNEEHPDSF 2283  
Qy 158 TAP-----GHSTFSFKALQ-PGLYVYHCAVAPVGMHIANMYGLILVEPKGSLPKVDXE 211  
Db 2284 VVPVAPSGDARRLTVSSSLQESGLKVNQAPASFAVSLNKAIDAKVHSPSGL-----SE 2339  
Qy 212 YVVMGQDFTKGYGQGLQPFDMEKAIREDAEYVV---FNGSVGLTGBNLLKAKVGET 268  
Db 2340 CYVTEID---QDKIAVRFIIP-----RENGVTLLIDVFNGI--HIPG-SPFKIRVGP 2385  
Qy 269 VALFVNGNG-PNLTSFHVIGEIFDKVHFEGKGEN---HNIQTLLIPAGGAIT----- 319  
Db 2386 ---CHGGDPGLVSAV-----GAGLEGVTVGNPAFAFVNTSAGAGALSVITDGS 2432  
Qy 320 EKVVD-----VQGDVYL-----VDHAIIPAFNKALGLIKVEG-- 353  
Db 2433 KVMDCQCEBGRVYVTPMAPGSYLSIKYGGPYHIGSPFPA-----KVTGPRLL 2483

Qy 354 -ENNEIYSHKQTDVAVLPEGAPQADITQEPAPKTPAPANLOEQIKAG-----KATY 402  
Db 2484 VSNHSLH---ELSSVFAVDSLTATCAPOHGARGPGBADSKVAVANGLSKAVYQGXSF 2540  
Qy 403 DSNCA-----ACHQD---DGKVPNAPPLANSDYLNADHARAASIVANGLSKI 449  
Db 2541 TVDCSKAGNNMLVGVHGPRTPCEEILVYHGVSRYSVSYLLKDKGEYTLVVK---WGHE 2597  
Qy 450 TVNGQYESVMP 461  
Db 2598 HTPGSPYRVVVP 2609  
  
RESULT 30  
US-10-156-761-10292  
; Sequence 10292, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OKURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 10292  
; LENGTH: 482  
; TYPE: PRT  
; ORGANISM: Streptomyces avermitilis  
US-10-156-761-10292  
  
Query Match 4.1%; Score 106.5; DB 15; Length 482;  
Best Local Similarity 22.7%; Pred. No. 0.37;  
Matches 85; Conservative 46; Mismatches 143; Indels 101; Gaps 20;  
  
Qy 114 GQMITVREBGTIEVOFSNHPDSKMPHNVDFHAATGP---GGGAASFTAPGHT--STSEF 168  
Db 88 GPTMKWTSQDVTVLNLTNSLADT--TVHFQGAHIPPXMDGQPONAFAP-GETWSPTFTV 144  
Qy 169 KALQGLVYVHCAVAPVGMHIANMYGLILVE---PKGGLPK---VDKEYVVMQ----- 216  
Db 145 KDEAKTLWHPRALDTTLEQVTRGLAGMIYEDSDSAAALSTYGTDDIPILLQCLAAD 204  
Qy 217 --GDF-YTKGKYGEOGLQPFDMEKAIREDAEVVENGSVGLTGENALKAKVGETVRLFV 273  
Db 205 SSGDVKYDLTGLVLSGLSLPILCNCTNVDAITLALT-ATKSRTRRLNAPSDDIITVGR 263  
Qy 274 GNGGPNLTSFHVIGEIFDKVHFEGKGENHNIQTLLIPAGGAITTEFVNDVPGDVLVD 333  
Db 264 GDGG--TLTQI-----ATDQGYLTEAAEVESIRIVAGARA--EFMDL----- 302  
Qy 334 HAIFRAFNKALGLIKVEGENEIIYSHKQTDVAVLPEGAPQADITQEPAPKTPAPANLQ 393  
Db 303 -----SDAVTL-----QAVVT-----TG 315  
Qy 394 QIKAGATYDSNCACHQDPDGKVPNAPPLAN--SDYLNADHARAASIVANGLSG-KIT 450  
Db 316 WVRGSGTYDFLTV---PDASDTPDALPSSLNITTRVDTDF--AARTITLGGGASML 370  
Qy 451 VNGQYESVMPALAL 465  
Db 371 INGSA-GTTMAAMM 384



US-10-156-761-14797

Query Match 4.0%; Score 104.5; DB 15; Length 625;  
Best Local Similarity 20.0%; Pred. No. 0.85;  
Matches 111; Conservative 55; Mismatches 191; Indels 197; Gaps 28;

QY 19 MLSGGSNOADKRAQPKSSTVDAAKTANADNAASQEHQELFVIDAIVTHAEVPPVDR 78  
DB 22 ILGKATSPA--AAAPRVTPVVVDVDDYGADPTGRTD---STPVAALHAKSV----- 70  
QY 79 DHPAAVVMVMEVEKMRDLADGEVQFWTFGGQVQGMIRVREGPTIEVQFN----- 131  
DB 71 DHPVIVSVSKGT-----YQLYP-----BRAETRELMSNTVYAGDOR 106  
QY 132 HPDSKMPHNV-DFHAATGPGGAE-----ASFAPGHT-STFSFKALQPGLYVH 179  
DB 107 YRDKKIGLLVEDMHVDVTDGSAKLVHGLQTAFAISIRSTDVTQNFSDYAP--EVID 164  
QY 180 CAAVAPVGMHIANGYGLILVEPKGLPKVDKEYVWQGF-----YTKGYGEGLOP 232  
DB 165 ATVAATG--VTDGHAYRVLKIIPAGSPRYVNGTHITWLGESTPATGQPYWGVGLQYTQI 222  
QY 223 FDMKALIEDAEVYVFNQSVGALTGENALKAVGEIVRLFVNGG----- 277  
DB 223 HDPEAQRTWRGDNPLFN-DVAAVTDLGGRRIRIDYTTARPADAGIVYQMLIERTEPGA 281  
QY 278 -----PNLT-----SSFVIGEI-----FDKVHFEKGKGNHNIQTLLIPAGGAI 318  
DB 282 FWEKSNVITMRSMAYILOSFGVVGQFSENISIDKVNFPD-----PNSGRST 329  
QY 319 TEFKVDVPGDYLVDHAIFFAFNKAGALGLK--VEGENHEIYSHKQTDAYVLPBGAPOA 376  
DB 330 ASF-----ADFQMGCV-----KGVKSITRSLFDGPHDPINIH---GTYL----- 367  
QY 377 IDTQAPKTPRANIOEQIKAGKATYDSNCAACHQPDGCVNAP-----P 423  
DB 368 -----EVGKPGPSTL-----TLAYKHPQTAGFPQAFPGDEVEPATKRTWTP 409  
QY 424 LANSDYLNADHRAASIVANGSGK-----ITVNGQYESV--MP 461  
DB 410 LADA-----HAQVYAV--DGPSGMDHTKPLTTMTVTTPRPVAGVETGTVVENTATP 461  
QY 462 AIALSDQOIANYIT 475  
DB 462 SVVISGNVFRNVPT 475

RESULT 34

US-10-137-870-234  
; Sequence 234, Application US/10137870  
; Publication No. US20030138883A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Beresini, Maureen  
; APPLICANT: DeForge, Laura  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Filvarioff, Ellen  
; APPLICANT: Gao, Wei-Qiang  
; APPLICANT: Gerritsen, Mary E.  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Sherwood, Steven  
; APPLICANT: Smith, Victoria  
; APPLICANT: Stewart, Timothy A.  
; APPLICANT: Tumas, Daniel  
; APPLICANT: Watanabe, Colin K  
; APPLICANT: Wood, William  
; APPLICANT: Zhang, Zemin  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; FILE REFERENCE: P3330R1C15  
; CURRENT APPLICATION NUMBER: US/10/137,870

; CURRENT FILING DATE: 2002-05-03  
; Prior Application removed - See Palm or File Wrapper  
; NUMBER OF SEQ ID NOS: 550  
; SEQ ID NO 234  
; LENGTH: 1160  
; TYPE: PRT  
; ORGANISM: Homo Sapien  
US-10-137-870-234

Query Match 4.0%; Score 104; DB 12; Length 1160;  
Best Local Similarity 19.7%; Pred. No. 2.5; Indels 106; Gaps 20;  
Matches 69; Conservative 58; Mismatches 116;

QY 102 EYQFWTFGQVY-----GQMRVREGPTIEVQSNHPSKMPHNVDFHAATG 148  
DB 791 EYTDGTF--RIPRPTGPEHGLIGLPIKGVGILVYFNK--NASPSYVH-- 842  
QY 149 PGGABASPTA-----PGHTSFSFKALQ--PG-----LYYHCAVAPVGMHIAN 192  
DB 843 ---GVLESTVWPLAEPGEVVTYQMNIPERSGPGPNDSACVSWIYSAVDPI-KDMVSG 898  
QY 193 MYGLI-----LVEPKGLPKVDKEYVW-----QSDPYTKGYGEGLOPFDMKAR 240  
DB 899 LVGPLAIQCKGLEPHGRSDMDREFALLPLIDENKSMYLEENVATHSQ--DPGSLN 956  
QY 241 EDAAEVVFNQSVGALTGE-----NALKAVGETVRLFVNGGPNL---TSFHVIGELPD 292  
DB 957 QDETLEEN-KHNAINGLXANLRGLTMYQSERVAMWMLAMQGDVLDLTHIHASFLX- 1014  
QY 293 KVHFEKGKGNHNIQTLLIPAGGAIITEFKVDVPGDYL---VDH-----AIFRAENK 342  
DB 1015 -----RNGENVRADVDLFPCTFEVEMVASNPCTWLMHCVTDHVHAGMETLFTVFSR 1068  
QY 343 GALTGLKXVEGENHEIYSHKQTDAYVLPBGAPOALDTQAPKTPRANIOE 393  
DB 1069 -----TEHL-----SPLTVITKETEKP--PROIEE 1092

RESULT 35

US-10-140-018-234  
; Sequence 234, Application US/10140018  
; Publication No. US20030138885A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Beresini, Maureen  
; APPLICANT: DeForge, Laura  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Filvarioff, Ellen  
; APPLICANT: Gao, Wei-Qiang  
; APPLICANT: Gerritsen, Mary E.  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Sherwood, Steven  
; APPLICANT: Smith, Victoria  
; APPLICANT: Stewart, Timothy A.  
; APPLICANT: Tumas, Daniel  
; APPLICANT: Watanabe, Colin K  
; APPLICANT: Wood, William  
; APPLICANT: Zhang, Zemin  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; FILE REFERENCE: P3330R1C158  
; CURRENT APPLICATION NUMBER: US/10/140,018  
; CURRENT FILING DATE: 2002-05-06  
; Prior Application removed - See Palm or File Wrapper  
; NUMBER OF SEQ ID NOS: 550  
; SEQ ID NO 234  
; LENGTH: 1160  
; TYPE: PRT  
; ORGANISM: Homo Sapien  
US-10-140-018-234

Query Match	4.0%	Score 104;	DB 12;	Length 1160;
Best Local Similarity	19.7%	Pred. No. 2.5;	Mismatches 116;	Indels 108;
Matches	69;	Conservative	58;	Gaps 20;
Qy	102	EYQFWTFCGGVDP-----	GMQIRVREGDTTIEVQESNHDSKMPHNVFHMATG	148
		:   :	:   :	
Db	791	EYTDQTF--RIRPRATGPEHNGILGPIIKGEVGDILVVFKN--	NASRPYSVAH-----	842
		:   :	:   :	
Qy	149	PGCGAASFTT-----	FGHTSTSEFKLQ--PG--	
		:   :	:   :	
Db	843	--GVLESTTWPPLAEBGEVVTYQMNIPERSGPGPNDSACVSYVASVDPI--	KDMSG 898	
		:   :	:   :	
Qy	193	MYGLI-----	LVEBKEGLPKVDKDYVYM-----	QSDPYTKKKKGEGGLQPFMEKXIR 240
		:   :	:   :	
Db	899	LVGFLATIQCKGILBHGGRSDMDREFALLFLIPDENKSWYLEBNVATHGSC--	DPGSLN 956	
		:   :	:   :	
Qy	241	EDAEVVENFGSVCALTGE-----	NALKAKVGETVLFEVGNCGPNL--	TSSFHYIGEIFD 292
		:   :	:   :	
Db	957	QDEFLFLEBN--KMHAINKGLYANLRCLTWYQGBERVAMWMLAMGQVDLHTIHFMHESFLY--	1014	
		:   :	:   :	
Qy	293	KVHEGGKGENHNHIIQTLLIPAGGAATFEKVDVPDGYVL-----	VDH-----	AIIPRAFNK 342
		:   :	:   :	
Db	1015	-----RNGENVRADVDLPGTFEVEVMVNSPQTWIMHGHVTDHVAAGMETLFTVESR	1068	
		:   :	:   :	
Qy	343	GALGILTKVEGEENHEIYSHKQTDVAUYLEPGARQALDIOEAKTPAPALQE	333	
		:   :	:   :	
Db	1069	-----TEHN--	SPLTATYETETKVP--	PROIE 1092
		:   :	:   :	

```

RESULT 36
US-10-140-021-234
Sequence 234, Application US/10140021
Publication No. US2003013886A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Beresini, Maureen
APPLICANT: DeForge, Laura
APPLICANT: Desnoyers, Luc
APPLICANT: Filvaroff, Ellen
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Sherwood, Steven
APPLICANT: Smith, Victoria
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K
APPLICANT: Wood, William
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P33081C167
CURRENT APPLICATION NUMBER: US/10/140,021
Pilot Application removed - See Palm or File Wrapper
NUMBER OF SEQ ID NOS: 550
SEQ ID NO 234
LENGTH: 1160
TYPE: PR1
ORGANISM: Homo Sapien
US-10-140-021-234

```

Query Match	4.0%	Score 104	DB 12	Length 1160
Best Local Similarity	19.7%	Pred. No. 2.5		
Matches	69	Conservative	58	Mismatches 116
				Indels 108
				Gaps 20
QY	102	EQYFPTFGQVP-----	GMIRAREGDTIEVQSNHDSKPHNVDPHATG	148
Db	791	EYTDPT--RIIPRTGPEHNLGILGPIIKGEVGLITLVFKN--NAPRPASVVAH----		842
QY	149	PGGGAENASFTL-----	PGHTSTFSFKALQ---PG-----	LYYHCAVAPVGMHIANG
				192

```

Db      ---GVLESTTWPLAEERGEVVTYQWNIPESSGCPDNDASCVMSTYYSAVDPI-KDMYSG  898
QY      193 MYGL-----LVPRKGLPKVDEEYVM-----QGDFTYKKGKYGEOGLQPFMEKXIR  240
Db      899 LVGPLAICQKILTEPHGGRSDMDPEFLLFLIPENKSWYLEENVATHGSO--PPGSINL  956
QY      241 EDAEYVFNQSVGALTGE-----NALKAKVGETVRLFEVNGGPNL---TSSFHVIGETIFD  292
Db      957 QDEFTLEESN-EMHAINKLYANLGLTMYOGERVAMWMLAMGDVDLHTTHFHAESEPLY- 1014
QY      293 KVFPEGGEGENHNITQTLLIPAGGAATEPKVDVGDVYL-----VDH-----AIFRAENK  342
Db      1015 -----RNGENYRADVDVDFEGTEFEVEMVMAISNGTILMHCHYVDHAGMETLFTYFSR  1068
QY      343 GALTGLKVEGENHEIYSHKOTDAVYLPEGAPOAIDPOEAKTPAPNLOE 393
Db      1069 -----TEHL-----SPLTVITKEKEV-PRDIEE 1092

```

```

RESULT 37
US-10-140-274-234
; Sequence 234, Application US/10140274
; Publication No. US20030143674A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Denoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerlitsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C161
; CURRENT APPLICATION NUMBER: US/10/140,274
; PRIORITY FILING DATE: 2002-05-06
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 234
; LENGTH: 1160
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-140-274-234

```

Query Match	4.0%*	Score 104;	DB 12;	Length 1160;	
Best Local Similarity	19.7%*	Pred. No. 2.5;			
Matches	69;	Conservative	116;	Indels 108; Gaps 20;	
QY	102	ELYQWTGGQVP-----	QGMIRVEDDTIEVQSNHFDKSMFNVDPHATG	148	
Db	791	EYDGTG--RIPRPTGPEEHLGILGLIKGEVDILTVFKN--NNSRPSTVHN----		842	
QY	149	PGGGAESAFTA-----	PGHTSTPSFPAQLQ---PG-----	LYVHCQAVAPVGHIANG	192
Db	843	---GVLESTTVPLAAPGEVTVTQWNIPEHSGGPNDSACVSIVYSAVDPI--KDMYSG	898		
QY	193	MYGLI-----	LVEPKKGLPKVDKEYVM-----	QGDFTYKGGYGQGGQAPFMEXAIR	240
Db	899	LVGEPLAICQXKILPHHGGKSDMDDEFLLPIIPENKSWIIEENVATHGSG--DPGSLNL	956		
QY	241	EDAEYVFNFGSVGALTGE---	NALKAVGETVRLFVNGGPNL---	TSSFHWIGELFD	292

```

Db      957 QDEFLEEN-KHNAINGKLYANRLGLTMYQSERVAMWMLAMQDQVDLHTIHFHASSFLY- 1014
Qy      293 KVHFEKGGENNIOTTLIPAGGAAITEFKVDPGDYVL-----VDH-----AIFRANFK 342
Db      1015 -----RNGENYRADVDLFPGTFEVEMVASNPGTWMHCHVDTHVHAGMETLFTVFSR 1068
Qy      343 GALGILKVBGEENHEIYSHKOTDAVYLPBGAPQALDTQCAPPTPAPANIOE 393
Db      1069 -----TEHL-----SPLVTYITKETEKVP-PRDIEE 1092

RESULT 38
US-10-140-471-234
; Sequence 234, Application US/10140471
; Publication No. US20030138887A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: Deforge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C163
; CURRENT FILING DATE: 2002-05-06
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 234
; LENGTH: 1160
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-140-471-234

Query Match      4.0%; Score 104; DB 12; Length 1160;
Best Local Similarity 19.7%; Pred. No. 2.5;
Matches 69; Conservative 58; Mismatches 116; Indels 108; Gaps 20;

Qy      102 EYQFWTFGGQV-----GOMIRVREGPTIEVQSNHPSDKMHPNVDFHATG 148
Db      791 EYTDGTF--RIPRPTGPEEHGILGPLIKGEVGDILTVFKN--NASRPSVHAH----- 842
Qy      149 PGGGAASFTA-----PGHSTSPFKALQ-----PG-----LYYHCAVAPVGNHANG 192
Db      843 ---GVLESTTWPLAEPGEVVTYQWNPERSGPGPNDSACVSWIYSAVDPI-KDMYSG 898
Qy      193 MYGLI-----LVEPKEGLPKVDKEYVW-----QGDFYTKGYGEGQLQPFDMKAIR 240
Db      899 LVGPLAIQOKGILPEHGSRSDMDREFALLFLFDENKSWYLEENAVATGSGQ--DGSINL 956
Qy      241 EDAEVVFNVSVALTGE-----NALKAKVGETVRLFVNGGPNL---TSSPHVIGELFD 292
Db      957 QDEFLEEN-KHNAINGKLYANRLGLTMYQSERVAMWMLAMQDQVDLHTIHFHASSFLY- 1014
Qy      293 KVHFEKGGENNIOTTLIPAGGAAITEFKVDPGDYVL-----VDH-----AIFRANFK 342
Db      1015 -----RNGENYRADVDLFPGTFEVEMVASNPGTWMHCHVDTHVHAGMETLFTVFSR 1068
Qy      343 GALGILKVBGEENHEIYSHKOTDAVYLPBGAPQALDTQCAPPTPAPANIOE 393
Db      1069 -----TEHL-----SPLVTYITKETEKVP-PRDIEE 1092

```

```

RESULT 39
US-10-140-807-234
; Sequence 234, Application US/10140807
; Publication No. US20030134354A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: Deforge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C174
; CURRENT FILING DATE: 2002-05-07
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 234
; LENGTH: 1160
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-140-807-234

Query Match      4.0%; Score 104; DB 12; Length 1160;
Best Local Similarity 19.7%; Pred. No. 2.5;
Matches 69; Conservative 58; Mismatches 116; Indels 108; Gaps 20;

Qy      102 EYQFWTFGGQV-----GOMIRVREGPTIEVQSNHPSDKMHPNVDFHATG 148
Db      791 EYTDGTF--RIPRPTGPEEHGILGPLIKGEVGDILTVFKN--NASRPSVHAH----- 842
Qy      149 PGGGAASFTA-----PGHSTSPFKALQ-----PG-----LYYHCAVAPVGNHANG 192
Db      843 ---GVLESTTWPLAEPGEVVTYQWNPERSGPGPNDSACVSWIYSAVDPI-KDMYSG 898
Qy      193 MYGLI-----LVEPKEGLPKVDKEYVW-----QGDFYTKGYGEGQLQPFDMKAIR 240
Db      899 LVGPLAIQOKGILPEHGSRSDMDREFALLFLFDENKSWYLEENAVATGSGQ--DGSINL 956
Qy      241 EDAEVVFNVSVALTGE-----NALKAKVGETVRLFVNGGPNL---TSSPHVIGELFD 292
Db      957 QDEFLEEN-KHNAINGKLYANRLGLTMYQSERVAMWMLAMQDQVDLHTIHFHASSFLY- 1014
Qy      293 KVHFEKGGENNIOTTLIPAGGAAITEFKVDPGDYVL-----VDH-----AIFRANFK 342
Db      1015 -----RNGENYRADVDLFPGTFEVEMVASNPGTWMHCHVDTHVHAGMETLFTVFSR 1068
Qy      343 GALGILKVBGEENHEIYSHKOTDAVYLPBGAPQALDTQCAPPTPAPANIOE 393
Db      1069 -----TEHL-----SPLVTYITKETEKVP-PRDIEE 1092

RESULT 40
US-10-140-922-234
; Sequence 234, Application US/10140922
; Publication No. US20030138889A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen

```

APPLICANT: DeForge, Laura  
APPLICANT: Desnoyers, Luc  
APPLICANT: Filvaroff, Ellen  
APPLICANT: Gao, Wei-Qiang  
APPLICANT: Gerritsen, Mary E.  
APPLICANT: Goddard, Audrey  
APPLICANT: Godowski, Paul J.  
APPLICANT: Gurney, Austin L.  
APPLICANT: Sherwood, Steven  
APPLICANT: Smith, Victoria  
APPLICANT: Stewart, Timothy A.  
APPLICANT: Tuma, Daniel  
APPLICANT: Watanabe, Colin K  
APPLICANT: Wood, William  
APPLICANT: Zhang, Zemin  
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
FILE REFERENCE: P330R1C179  
CURRENT APPLICATION NUMBER: US/10/140,922  
CURRENT FILING DATE: 2002-05-07  
Prior Application removed - See Palm or File Wrapper  
NUMBER OF SEQ ID NOS: 550  
SEQ ID NO 234  
LENGTH: 1160  
TYPE: PRT  
ORGANISM: Homo Sapien  
US-10-140-922-234

Query Match 4.0%; Score 104; DB 12; Length 1160;  
Best Local Similarity 19.7%; Pred. No. 2.5; Matches 69; Conservative 58; Mismatches 116; Indels 108; Gaps 20;

QY 102 EYGFMTFGGVP-----GGMIRREGDTIEVOFSNHPDSKMPHNVDPHAATG 148  
DB 791 EYDGTFR-RPRPRGPEEHLGLPPLKGEVGDITVFKN--NSRPSVHAH---- 842  
QY 149 PGGAASFTA-----PGHSTFSFKALQ---PG-----LYYHCAVAPVGMHANG 192  
DB 843 ---GVLESTVWPLAEPGEVVTYQWNIPEKSGPNDACVSMIYSAVDPI-KDMYSG 898  
QY 193 MYGLI-----LYEPKGLPKVDKEYVM-----QGFYTKGKYGEGCLQPPDMKXIR 240  
DB 899 LVGPPLAICQKGLPEHGRSDMDREFALFLIPDENKSWYLEBNVATHSQ--DPGSINL 956  
QY 241 EDAYVVFNGSVGALTGE-----NALKAKVGETVRLPVNGGPNL---TSSFVIGEIFD 292  
DB 957 QDETFLBSN-KMAHINKLYANLRGLTYGGERVAMTLMAGQVDLHTIHFAESFLY- 1014  
QY 293 KVAFEGGKGENHNIQTLLIPAGGAATFEKVDVPGDYVL---VDH-----AIPRAFNR 342  
DB 1015 -----RNGENYRADVDLFPGTPEVEMVASNPGTWMHCHVDVHAGMETLFTVFSR 1068  
QY 343 GALGILKVEGGENHEIYSHKOTDAVYLPBGAPOAIDTQEAPKTPAPANLOE 393  
DB 1069 -----TEHL-----SPLTVITKETEKVP-PRDIEE 1092

Search completed: August 27, 2003, 18:34:16  
Job time : 33 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 27, 2003, 18:28:17 ; Search time 29 seconds

(without alignments)  
732.416 Million cell updates/sec

Title: US-10-088-045-2

Perfect score: 2626

Sequence: 1 MSKPLIKITLICALSALML.....NKGGLSADVAKKXKXEN 502

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	163.5	6.2	645	US-09-252-991A-25253	Sequence 25253, A
2	163	6.2	442	US-09-296-284-26	Sequence 26, App1
3	163	6.2	478	US-09-296-284-5	Sequence 5, App1
4	153	5.8	751	US-09-252-991A-25481	Sequence 25481, A
5	150	5.7	444	US-09-252-991A-17336	Sequence 17336, A
6	150	5.7	712	US-09-252-991A-21481	Sequence 21481, A
7	147.5	5.6	658	US-09-252-991A-24910	Sequence 24910, A
8	146	5.6	468	US-07-985-458-4	Sequence 4, App1
9	146	5.6	593	US-09-252-991A-30673	Sequence 30673, A
10	145	5.5	559	US-08-030-096-6	Sequence 6, App1
11	141	5.4	529	US-08-172-331B-14	Sequence 14, App1
12	141	5.4	529	US-08-706-037-25	Sequence 25, App1
13	141	5.4	529	US-09-005-397-25	Sequence 25, App1
14	141	5.4	529	US-09-032-315-5	Sequence 5, App1
15	141	5.4	529	US-08-993-318A-5	Sequence 5, App1
16	141	5.4	529	US-09-399-886-5	Sequence 5, App1
17	141	5.4	529	US-09-396-260-5	Sequence 5, App1
18	141	5.4	529	US-09-576-281-5	Sequence 5, App1
19	136.5	5.2	326	US-09-615-192A-291	Sequence 291, App
20	136	5.2	320	US-09-252-991A-23061	Sequence 23061, A
21	135	5.1	575	US-09-032-315-8	Sequence 8, App1
22	135	5.1	575	US-08-993-318A-8	Sequence 8, App1
23	135	5.1	575	US-09-399-886-8	Sequence 8, App1
24	135	5.1	575	US-09-396-260-8	Sequence 8, App1
25	135	5.1	575	US-09-576-281-8	Sequence 8, App1
26	133.5	5.1	576	US-08-172-331B-2	Sequence 2, App1
27	132	5.0	671	US-09-328-352-6003	Sequence 6003, Ap

28	131.5	5.0	635	US-09-328-352-8029	Sequence 8029, Ap
29	130	5.0	283	US-08-746-111-5	Sequence 5, App1
30	129	4.9	540	US-08-687-580B-7	Sequence 7, App1
31	125.5	4.8	310	US-09-252-991A-23088	Sequence 23088, A
32	124	4.7	572	US-09-032-315-7	Sequence 7, App1
33	124	4.7	572	US-08-993-318A-7	Sequence 7, App1
34	124	4.7	572	US-09-399-886-7	Sequence 7, App1
35	124	4.7	572	US-09-396-260-7	Sequence 7, App1
36	124	4.7	572	US-09-576-281-7	Sequence 7, App1
37	121.5	4.6	539	US-09-032-315-1	Sequence 1, App1
38	121.5	4.6	539	US-08-993-318A-1	Sequence 1, App1
39	121.5	4.6	539	US-08-689-421-27	Sequence 27, App1
40	121.5	4.6	539	US-09-399-886-1	Sequence 1, App1
41	121.5	4.6	539	US-09-396-260-1	Sequence 1, App1
42	121.5	4.6	539	US-09-389-528-27	Sequence 27, App1
43	121.5	4.6	539	US-09-181-827A-27	Sequence 27, App1
44	121.5	4.6	539	US-09-576-281-1	Sequence 1, App1
45	120.5	4.6	2343	US-09-324-867-2	Sequence 2, App1

## ALIGNMENTS

RESULT 1  
US-09-252-991A-25253  
Sequence 25253, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 25253  
LENGTH: 645  
TYPE: PRT  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-25253

Query Match	6.2%	Score 163.5;	DB 4;	Length 645;
Best Local Similarity	21.8%	Pred. No. 2.1e-06;		
Matches	93;	Conservative 52;	Mismatches 141;	Gaps 21;
QY	19	MTSGCSNQADKA-AOPKSSSTVDAAAKTANADNAASQHGELPVI-----DAIVTHAPEV 72		
DB	41	VLAAGSTPFAKFAAQGGSAT-----RGAPVLKGTFFDVIASPV- 81		
QY	73	PPPVDRHDAKVVVMEVEKVMRLADGVEYQFWTFGGVPCQMIRVREGDTIEVQSNH 132		
DB	82	-----NFGKRGVAT-----TINGSLPAPTLRWREGDFTIRVTN- 116		
QY	133	PSKMPHNDVFAAAGP---GGGAASFT--APGHSTSPFALQGLVYHCAVAPVCM 187		
DB	117	KLEKRTSIIHMGIIILPYOMDVGPISEFGIAPGETFYFRFVQSGSYWH---SHSGF 172		
QY	188	HIANGVYGLIPEKREGLPKV---DKEYVYM-----OGDFYT----- 221		
DB	173	QELTMYGAILIIDP-AGATITHADRDHVLIFSDMTDEDPMKFTFKKSQSDYNNVQPTV 231		
QY	222	-----KGYGEOGLQFDMEKAIREDAYEVNGSVGALTGENALKA 263		
DB	232	PFDFDASREGTAAAIIDKKMMNEMKMNPTDLADISATLIFYLA--NGVTPA--GWTALF 288		
QY	264	KXGETVRLPVGCGNLTSSPHVIGEIPKXVPEGKGNNHIIQTLIAPAGAAITFEFV 323		
DB	289	RGGERVRLMVGAGNTFYDVRIPLGLKLTIVHVDG-----V 324		

Oy	334	DVPGGVVLVDNHAI	FR	FNKG	-	ALGILKVEEENHEIYSHQOTAVVLP	REGAP	POAIPDQEA	382
Db	325	DV--	SPVTVDE--	FR	-	FGPGETVTVIVQPRDDATYTI	FAQAMDBTGY----	ARATLAVREG	375
Oy	383	PKTPAPA	389						
Db	376	LQAPVPA	382						

RESULT 2  
US-09-296-284-26  
; Sequence 26, Application US/09296284A

1 APPLICANT: Choi, Eun-Sung  
 2 APPLICANT: Rhee, Sang-Ki  
 3 APPLICANT: Lee, Eun-Hee  
 4 TITLE OF INVENTION: Glucanobacter Suboxydans Sorbitol Dehydrogenase, Genes  
 5 TITLE OF INVENTION: And Methods of Use Thereof  
 6 FILE REFERENCE: 1533.087000  
 7 CURRENT APPLICATION NUMBER: US/09/296,284A  
 8 CURRENT FILING DATE: 1999-04-22

```

; GC: 50.14%
; SEQ ID NO 26
; LENGTH: 442
; TYPE: PRT
; ORGANISM: Gluconobacter suboxydans
US-09-296-284-26

```

Query Match	6.2%	Score 163;	DB 3;	Length 442;
Best Local Similarity	37.3%;	Pred. NO. 1.2e-06;		
Matches 50;	Conservative 17;	Mismatches 57;	Indels 10;	Gaps 6

QY 369 LPEGAQALDTQEARPTTPAPANLQEQIKAGKATYDSNCAACGQPDGCGVPNAPPLNSD 428  
Db 270 VPEGKNLQ---QDDGKATALLIAGGKADGAEVYLHNCAICPMNDGCGVNRNFPPLGNP 326

QY 429 YLNDHARA-ASIVANGLSGKITVNGQYES-MPALA-LSDDOIANVITYTLNSFGNK 484  
327 VVITNDATSMANIVTFG--GLIPLPTNTAPSAAMGCRDHLSDQIADVNMERKSGNQ 384  
Db

```
QY      485 G-GQLSADDVAKAK 497
          |||||:|:
Db      385 APGTLASDIRKLR 398
```

RESULT 3  
US-09-296-284-5  
: Sequence 5, Application US/09296284A

1 GENERAL INFORMATION:  
2 APPLICANT: Choi, Eui-Sung  
3 APPLICANT: Rhee, Sang-Ki  
4 APPLICANT: Lee, Eun-Hee  
5 TITLE OF INVENTION: Glucanobacter Suboxydans Sorbitol Dehydrogenase, Genes  
6 TITLE OF INVENTION: And Methods of Use Thereof  
7 FILE REFERENCE: 1533.0870000  
8 CURRENT APPLICATION NUMBER: US/09/296,284A

```

; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 478
; TYPE: PRT
; ORGANISM: Gluconobacter suboxydans
US-09-296-284-5

```

Query Match 6.2%; Score 163; DB 3; Length 478;  
Best Local Similarity 37.3%; Pred. No. 1.4e-06;  
Matches 50; Conservative 17; Mismatches 57; Indels 10; Gaps 6

Db 306 VPEGKULG---QDDKATALLLEAGSKGDAAEVYLTHNCALICHNMIDGTVGVRMEPEPLAAGNP 362

Qy 429 YLNADHARA-AASIVANGSLGKTTVNGNYEYV-MPALA--LSDQILANVTYTLTNSGAK 484

Db 363 VVITDNTAMNIVTFG-GILPPTNTPASVAMPGFRDLISDQILDVNVFMKRSWGNQ 420

```

QY      485 G-GQLSADDVAKAK 497
          |||||:|:
Db      421 APGTLASDIRKLR 434

```

RESULT 4  
US-09-252-991A-25481

```

: GENERAL INFORMATION:
: APPLICANT: Marc J. Rubenfield et al.
: TITLE OF INVENTION: NUCLEIC ACID SEQUENCES RELATING TO PSEUDOMONAS
: TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
: FILE REFERENCE: 107196.136
: CURRENT APPLICATION NUMBER: US/09/252.991A

```

```

;
; ERROR APPLICATION NUMBER:  US 60/074,760
;
; PRIOR FILING DATE: 1998-02-18
;
; PRIOR APPLICATION NUMBER:  US 60/094,199
;
; PRIOR FILING DATE: 1998-07-27
;
; NUMBER OF SEQ ID NOS: 33142
;
; SEQ ID NO 25481

```

```

; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; FEATURE:

```

LOCATION: (339)  
OTHER INFORMATION: Identity of amino acid at the above locations are unknown  
US-09-252-991A-25481

Query Match	5.8%;	Score 153;	DB 4;	Length 751;
Best Local Similarity	35.5%;	Pred. No. 2.5e-05;		
Matches 38;	Conservative 15;	Mismatches 48;	Indels 6;	Gaps 4

Qy 396 KAGKATYDUSNCACHQDPDGKGVNPAFPPLA-NSDYLNADHARAASIVANGSGKITVNGN 454  
| | | | | : | | | | | :  
Db 626 KTGALTYDNCGACHRTDSKGVARVPALAGNPVTVGSDPTSLVHVLKG-GTLPATHQ 683

```
QY      455 QYES-VMPAIA--LSDQQIANVITYTLNSFGNKGCGLSADDVAKAKK 4988
```

RESULT 5  
US-09-252-991A-17336  
Sequence 17336. Application US/09252991A

APPLICANT: Marc J. Rubenfield et al.  
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOWOMAS  
 TITLE OF INVENTION: AEBUIGNOSA FOR DIAGNOSTICS AND THERAPEUTICS  
 FILE REFERENCE: 107196.136  
 CURRENT APPLICATION NUMBER: US/09/252.991A  
 CURRENT FILING DATE: 1999-02-18  
 PRIOR APPLICATION NUMBER: US 60/074.788  
 PRIOR FILING DATE: 1998-02-18  
 PRIOR APPLICATION NUMBER: US 60/094.190  
 PRIOR FILING DATE: 1998-07-27  
 NUMBER OF SEQ ID NOS: 33142

```

; ORG: 1700
; LENGTH: 444
; TYPE: PRF
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-17336

```



Query Match 5.7%; Score 150; DB 4; Length 444;  
 Best Local Similarity 22.0%; Pred. NO. 2e-05;  
 Matches 63; Conservative 38; Mismatches 86; Indels 100; Gaps 11;

QY 265 VGETVLFVNGGPNLTSSPHVIGEIFDKVHFGGKGNNTIOT---TLIPAGGA----- 316  
 DB 205 LGQDVYFSLNATPQ-----DQIHNRQAKDENHLELVEDEPLVPGTKYRFLI 252  
 QY 317 -----AITEF-----KVDVPGDY-----VLVDHA 335  
 DB 253 TSSDVIHSMWVPAFAVKRDLIPGFVNEAMTKVDEPGYRQCACELCKDHFMPIYVD-- 310  
 QY 336 IFPAFNKAGALGILKVEEENHEIYSHKQTDVYLPEGAPQIDTQEPKTPAPANIQEOI 395  
 DB 311 -----VKPRAEFPQWLAKRKE-EAAKYKE-----LTSKEWTK-----ELV 345  
 QY 396 KAGKATYDSNCACHQPDGKGVNAPPLANSDYLNADHARAASIYANGLSGKITVNGNQ 455  
 DB 346 ARGDKVYHTTCAACHQAEQGMPEPALKGSKIYTGPKKHLELVFNVGPG----- 397  
 QY 456 YESVMPAIA--LSDQOIANVITYTLNSFGNGKGLSADVAKAKKTK 500  
 DB 398 --TAMAAFGKQINLEVDLANVITIERANMGNDGDMTPKRVAVAKK 442

RESULT 6  
 US-09-252-991A-21481  
 ; Sequence 21481, Application US/09252991A  
 ; Patent No. 6551795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ; FILE REFERENCE: 107196.136  
 ; CURRENT APPLICATION NUMBER: US/09/252.991A  
 ; PRIOR FILING DATE: 1999-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/074.788  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/094.190  
 ; PRIOR FILING DATE: 1998-07-27  
 ; NUMBER OF SEQ ID NOS: 33142  
 ; SEQ ID NO 21481  
 ; LENGTH: 712  
 ; TYPE: PR1  
 ; ORGANISM: Pseudomonas aeruginosa  
 US-09-252-991A-21481

Query Match 5.7%; Score 150; DB 4; Length 712;  
 Best Local Similarity 26.3%; Pred. NO. 4.4e-05;  
 Matches 46; Conservative 28; Mismatches 73; Indels 28; Gaps 6;

QY 347 ILKVEEENHEIYSHKQTDVY-----LPEGAPQIDTQEPKTP 386  
 DB 281 LKLT--GQNRPAVYSGMTVDVHSHSTQHSDDLLAIAYSKSLPAGKDDLPMPDSERPILA 339  
 QY 387 APANLQEOIKAGKATYDSNCAACHQPDGKGVNAPPLANSDYLNADHARAASIYANGLS 446  
 DB 340 APVDLVS--SRGLGTAQFSCDHRKDGSGVPCGFPLAANGPYVAS--ANPSTLHLHTLT 395  
 QY 447 GKITVNGNOYEV--MPAIA--LSDQOIANVITYTLNSFGNGKGLSADVAKAKK 498  
 DB 396 GWTATATATSHRYVTMGFAQLDRELAIFLISFVRSSWNGQSSIDAGQYKLRQ 450

RESULT 7  
 US-09-252-991A-24910  
 ; Sequence 24910, Application US/09252991A  
 ; Patent No. 6551795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ; FILE REFERENCE: 107196.136

CURRENT APPLICATION NUMBER: US/09/252.991A  
 ; CURRENT FILING DATE: 1999-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/074.788  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/094.190  
 ; PRIOR FILING DATE: 1998-07-27  
 ; NUMBER OF SEQ ID NOS: 33142  
 ; SEQ ID NO 24910  
 ; LENGTH: 658  
 ; TYPE: PR1  
 ; ORGANISM: Pseudomonas aeruginosa  
 US-09-252-991A-24910

Query Match 5.6%; Score 147.5; DB 4; Length 658;  
 Best Local Similarity 21.2%; Pred. NO. 6.6e-05;  
 Matches 85; Conservative 56; Mismatches 150; Indels 109; Gaps 20;

QY 107 TREGQVPGQIMIRYREEDTTEVQFSNH--PDSK-----MPHNVDPHAATGPGGGAASR 157  
 DB 110 TTNGSLPGPTLWRREGDNTLVRNRLAEDTSIHMHGILPANMD-----GVPLSP 161  
 QY 158 --TAPGHTSPSKALQPLGYVYHCAVAPVGMHIANGMTGLIVEPKGLP-KVDKEYVY 214  
 DB 162 EGIAPGGLYERPKVRQNGITWTH--SHSGLOEQAGVIGALVIDAKREPPPSYDNDYV 218  
 QY 215 MGDFFYTKGKYGEOGLQPFPMDEKAIREDAYVVF-----NGSVGALTGEN- 259  
 DB 219 LSTDWSE-----KQRIALKLKKQSDYVNFHRTYGDFFIDVVSANGMAATLDRKM 270  
 QY 260 --ALK-----AKVGTVLFVNGNGP---NLTSFHVIGEIFDKVHFBGKGNNTIOT 308  
 DB 271 MAEMKSPPTDLADVSQYTYTLNGQPPDGMNGLFR-----PEKRLRFR 316  
 QY 309 TLIPAGGAATPEKRVDPG-----DYLVDH--AIFPAFNKAGALGILKVEG 352  
 DB 317 ---VNASMSYFDVRIPELKMNTVVAADQHVPEVSDERLAVATYDV---IVPPG 368  
 QY 353 EENHEIYSHKQTDVYLPEGAPQIDTQEP--KTPAPANIQEOIKAGKATYDSNCAACHQ 411  
 DB 369 ERAYTLFAGSMDSRGV--ARGTLALAGLSAPVTPPRPLIGMDMGMDGMD-----HG 421  
 QY 412 PDGKGVNPA-----PPLANSDYLNADHARAASIYANGLS 445  
 DB 422 AMGHGATRPASEMDHSKMSGMDKMGDHSKXAGMDMNGM 461

RESULT 8  
 US-07-985-458-4  
 ; Sequence 4, Application US/07985458  
 ; Patent No. 5344777  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Tamaki, Toshimi;  
 ; APPLICANT: Takemura, Hitoshi;  
 ; APPLICANT: Takemura, Kenji;  
 ; APPLICANT: Fukaya, Masahito;  
 ; APPLICANT: Okumura, Hajime and  
 ; APPLICANT: Kawamura, Yoshiya  
 ; TITLE OF INVENTION: Structural Gene of Membrane-Bound  
 ; TITLE OF INVENTION: Alcohol Dehydrogenase Complex, Plasmid  
 ; TITLE OF INVENTION: Containing The Same And Transformed Acetic Acid  
 ; NUMBER OF SEQUENCES: 4  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Frischauf, Holtz, Goodman & Woodward, P.C.  
 ; STREET: 600 Third Avenue  
 ; CITY: New York  
 ; STATE: New York  
 ; COUNTRY: USA  
 ; ZIP: 10016-2088  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: 3.5 inch, 0.72 mb  
 ; COMPUTER: IBM PC compatible (NEC PC-9801 ES)  
 ; OPERATING SYSTEM: MS DOS

SOFTWARE: ASCII Form  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/985,458  
FILING DATE: 19921203  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/658,221  
FILING DATE: 20-FEB-1991  
APPLICATION NUMBER: 73440/1990  
FILING DATE: 26-MAR-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Goodman, Herbert  
REGISTRATION NUMBER: 17081  
REFERENCE/DOCKET NUMBER: 910134/HG  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212)972-1400  
TELEFAX: (212)370-1622  
TELEX: 236268  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 468 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
FEATURE:  
NAME/KEY: MATURE PEPTIDE  
LOCATION: 24 to 468  
IDENTIFICATION METHOD: similarity of other signal  
IDENTIFICATION METHOD: sequences  
ORIGINAL SOURCE:  
ORGANISM: Acetobacter alioacetigenes  
STRAIN: MH-24  
PUBLICATION INFORMATION:  
AUTHORS: Tamaki, Toshimi;  
AUTHORS: Fukaya, Masahiro;  
AUTHORS: Takemura, Hiroshi;  
AUTHORS: Takemura, Kenji;  
AUTHORS: Okumura, Hajime;  
AUTHORS: Kawamura, Yoshiya;  
AUTHORS: Nishiyama, Makoto;  
AUTHORS: Horiouchi, Sueharu and  
AUTHORS: Bepu, Teruhiko  
TITLE: Cloning and Sequencing of the Gene Cluster  
TITLE: Encoding Two Subunits of Membrane-Bound  
TITLE: Alcohol Dehydrogenase from Acetobacter  
TITLE: polyoxogenes  
JOURNAL: Biochimica et Biophysica Acta.  
VOLUME: 1088  
PAGES: 292-300  
DATE: 1991  
US-07-985-458-4

Query Match 5.6%; Score 146; DB 1; Length 468;  
Best Local Similarity 25.2%; Pred. No. 5.2e-05;  
Matches 84; Conservative 50; Mismatches 127; Indels 72; Gaps 19;

184 PVGHNINAGVGLL-VERKELPKVDKEYVVGMDPYTKGKGEOGL-----QPEDM-EK 237  
149 PLSNRMLGIMRMFSSPKDFTPRAGTDEIARGLVLTGP-CHCAGCTPRGFAMQEK 207  
238 AIRDEAVVFNSSVGLTGENALKAKVGETVR-----LFVNGGPNLTSSFHVIGEIFDK 293  
208 AL--DA-----GGPDLSSGAPRIDNVNAPSLRNDPVVGGRWSEDDIYFLKSGRIDHS 260  
294 VHEEGGGENHNITQTLIPAGGAAITPKVDVPEDDYLVN--HAIFPAFKALGILKVE 351  
261 AVF-GGMD-----VVAMSTQYFTDDLAIAK-YLKSJRPVPSQ 299  
352 GEENHEIYSHKQTDVAVLPREGARQIDTQEAQKTPRANIOEQIKAKATYDSNCAACHQ 411  
300 GN-----YTDPSFANMLASG-----NTASVP-----GADTYVEKALCHR 335  
412 PDGKGVNAPPLANSDYLANADHARA-ASIVANGLSKITVNGQYESV-MPAI--ALSD 467

336 NDGGVARNMPPLAGPVVVTENPTSLVNVIAHG--GVLPSSWABSAVAMPQYSKLSA 393  
468 QQIANVITYTLNSFGNKG--QGLSADDAVAKAKT 499  
394 QQIADVNFRTISWGNKAPGTVAADVTGLRDT 426

RESULT 9  
US-09-252-991A-30673  
Sequence 30673; Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubinfeld et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 30673  
LENGTH: 593  
TYPE: PRT  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-30673

Query Match 5.6%; Score 146; DB 4; Length 593;  
Best Local Similarity 22.1%; Pred. No. 7.7e-05;  
Matches 116; Conservative 61; Mismatches 219; Indels 126; Gaps 25;

68 HADEVPPVDRDHPRAVTVVVMETVEKMRADCEVQFTFGQVGV--QMTREGEQT 124  
92 HPRABDDPAVRRPVLVPLV-----LRAGLRORRRRAGGKSPAPADQORAGACDR 143  
125 IEVQFSNHPDSKMPHNVDFH-----AATGPG-GGAASFPAPGHTSFSPKALOP----- 173  
144 GQPRSHLPLHRIALLQHNPPRARTTRPGGGLRKRTLSC--LALALALSPFLASAA 201  
174 -----GLVYH-----CAVAPVGNHANGM-----YGLILVEPK-----GLP 206  
202 GADALVKRGEYLARADACHTAGAPFAGGLIQSPFTIYGTNTITPKHEGIGAY 261  
207 KVDKEYVNMGGDYTKGKGEOG---LQPFMEKAI-REDAEYV----- 247  
262 SADEFFPAL-----TEGKRKDGAVLYPAMPYTSYHLIEREDADAIYALMAQEPRIARPAP 316  
248 -----FNGSVGALIGENALKAKVGETVRLFVNGGPNLTSSFHVIGEIFDKVHREGK 300  
317 QTSLSFPFVNRNG-LAGMNL-----YKGSVRLQPEBGRSEAMKRGQYVEVL-----GHC 366  
301 GENHNITQTLIPAGGAITEFKVD-----VPGD-----YVLVDHAIFFAFNKG 343  
367 GEHTNRN-----LAGLEDDKRLSGGLNGYLAPSLRPDLARGTQGLDLSFLHGMS 422  
344 ALGIL-----KVEGENHEIYSHKQTDVY-LPREGARQIDTQEAQKTPRANIOEQIK 396  
423 AQSMEFENFPPVNHSTQHLDDSLAAMATYLLGDPQPPAKVQALPE-----AQLNDSG 478  
397 AGKATYDSNCAACHQPDGKGVN-APPPLANSDYLANADHARAASIVANGLSKITVNGNO 455  
479 RGRQQLNVCAGHGSGGKPRHIAVAMNGNTTLRLQDRPNLIRVIEDGIVBO-QRTGFE 537  
456 YESVMPAIA--LSDOQIANVITYTLNSFGNKGGLSADDAVAKAK 497  
538 RMQPMGFGAKLDDBDLTLLNLYLRTWGLGFQDLSRQVQALX 581

RESULT 10  
US-08-030-096-6

; Sequence 6, Application US/08030096  
; Patent No. 5426041  
; GENERAL INFORMATION:  
; APPLICANT: Fabijanski, Steven F.  
; APPLICANT: Arriason, Paul G.  
; TITLE OF INVENTION: BINARY CRYPTOCYTOTOXIC METHOD OF HYBRID  
; TITLE OF INVENTION: SEED PRODUCTION  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 3000 K Street, N.W., Suite 500  
; CITY: Washington, D.C.  
; COUNTRY: USA  
; ZIP: 20007-5109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/030,096  
; FILING DATE: 22-MAR-1993  
; CLASSIFICATION: 800  
; PRIORITY APPLICATION DATA:  
; APPLICATION NUMBER: US 07/556,917  
; FILING DATE: 20-JUL-1990  
; PRIORITY APPLICATION DATA:  
; APPLICATION NUMBER: PCT/CA91/00255  
; FILING DATE: 22-JUL-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BENT, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 33229/164/PIHI  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202)672-5300  
; TELEFAX: (202)672-5399  
; TELEX: 904136  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 559 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-030-096-6

Query Match 5.5%; Score 145; DB 1; Length 559;  
Best Local Similarity 26.1%; Pred. No. 8.7e-05;  
Matches 71; Conservative 37; Mismatches 132; Indels 32; Gaps 12;

QY 100 GVEYQFWTFGGQVPGQMIRVREGDTLEVOFSNHPD-----SKMPHNVDFHAATGPGGG 152  
DB 43 GVPQGVILINGQFPGNNINSTNNNNIIVFNNLDEPFLITNNGIGHKKNQMDGTFPGIM 102  
QY 153 AEAATAPGHTSTFSGKAL-QPGLYYHCAVAVGMIHANGYGLILVEPKGLP-----K 207  
DB 103 CP---IMPETNYTHYHPOKDOIISYFY---PTTGMRAGAGYGGYGRVNSRLIIPVPYAD 156  
QY 208 VDEKEYVMGDFYTKGKGEQGLQPF-DNEKAI-REDAEYVVPNGSVGLTGENA--LKA 263  
DB 157 PEDDTVLIGDMYTK--SHTQIKKPLDGRITIGRPGD--IYINGSGKGGSDALEFTL 211  
QY 264 KVGETVRLFEVNGGPNLTSSPHVIGEIFDKVHFEKGGENHNIOITLLPAG--GAITE 320  
DB 212 KKGKTYRVAICVGVTSINFRIOHNMKMLVEMESHVLONDYDLSLVAVGCGFTIVTA 271  
QY 321 FKVDVPGDVLVDHAIERAFNKGALGILKEVG 352  
DB 272 NQ--EPKDYVYVASSRFLKTVITTTGLRYEG 301

RESULT 11  
US-08-172-331B-14  
; Sequence 14, Application US/08172331B

; Patent No. 5480801  
; GENERAL INFORMATION:  
; APPLICANT: Wahlethner, Jill A.  
; APPLICANT: Christensen, Bjoern E.  
; APPLICANT: Schneider, Ralle  
; TITLE OF INVENTION: PURIFIED PH NEUTRAL LACCASES AND NUCLEIC  
; TITLE OF INVENTION: ACIDS ENCODING SAME  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: No. 5480801 of No. 5480801 of No. 5480801 of America, Inc.  
; STREET: 405 Lexington Avenue  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10174-6401  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/172,331B  
; FILING DATE: 22-DEC-1993  
; CLASSIFICATION: 435  
; PRIORITY APPLICATION DATA:  
; APPLICATION NUMBER: US 08/122,230  
; FILING DATE: 17-SEP-1993  
; PRIORITY APPLICATION DATA:  
; APPLICATION NUMBER: US 08/122,827  
; FILING DATE: 17-SEP-1993  
; PRIORITY APPLICATION DATA:  
; APPLICATION NUMBER: US 08/162,827  
; FILING DATE: 03-DEC-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Lowney Dr., Karen A.  
; REGISTRATION NUMBER: 31,274  
; REFERENCE/DOCKET NUMBER: 4052.020-US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-867-0123  
; TELEFAX: 212-878-9655  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 529 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-172-331B-14

Query Match 5.4%; Score 141; DB 1; Length 529;  
Best Local Similarity 20.4%; Pred. No. 0.00019;  
Matches 103; Conservative 60; Mismatches 216; Indels 126; Gaps 20;

QY 59 LPVIDAIVTHADPEVPPVDRDHPAKVVK--METVEKVRMLADGVEYQFWTFGGQVPGQM 116  
DB 8 LPLAAVST-----PAPAFAVNYKFDIKNVAVADGFGORSIVSVNGVLPCTL 54  
QY 117 IVRREGDTLEVOFSNH---PDSKMPHNVDFH---AATPGGGAESSTFAP-----GHS 164  
DB 55 ITANKGDTLRINVTQLODPSNRRAATTIHHGLFOATTADEDEGPAFTVTCPTAQNLSYTY 114  
QY 165 TFSFALPGLLVYCAVAPVGMHIANGYG-LILVEP---KEGLPKVDKEYVYVMOGDF 219  
DB 115 EIRPLNG-QGTGMWYH---AHLASQVVDGLRGLVYIDPRDPKRSRYDVDDASTVNLLEW 170  
QY 220 Y-----TKGKYGEGQ---IQPFMEKAIREDAYVVPNGSVGLTGENALKAKVGETV 269  
DB 171 YHTPAVLEKQMFSTNNNTALLSPVPDGLINGKGRYV-----GCPAVPRSVINVKRGKXY 225  
QY 270 RLFGVNGGNLTSFHVIGEIFDKVHFEKGGENHNIOITLLPAG--AAITEFKVD 324  
DB 226 RLRVINASAGISFTSY-----EGHSL--TVIEADGILHQPLAVSFOLY 268  
QY 325 VPGDVLVDHAIERAFNKGALGILKEVEGENHEIYSHKOTDAVYVPEGAPQADTQE--- 381

Db 269 AGORYSVIVEANOTANWIRAPMTVAGAGTNALDPTNVFAVLHGYGADNAPTTEQGS 328  
QY 382 -----AKTPAPANLQOIKAGKATYDS-----NCAACHQPGKGV 417  
Db 329 AIGTALVEENLHALINPGAPGASAPADVSLNLAIGRSTVDGILRFTFNRIKYEAP----- 383  
QY 418 PNAFPPLANSDYLNADHARAASIVANGLSGITVNGQYESVWPAIALSDQOIANVITYT 477  
Db 384 --SLPTL-----LKLANNASNDADFTNEHTIYLPHNKVIELNITGADHP 428  
QY 478 LNSFGKGGQLSADDAVAKAKTKPN 502  
Db 429 IHLHG-----VFIDVKSLSGTFN 447

## RESULT 12

US-08-706-037-25  
; Sequence 25, Application US/08706037  
; Patent No. 5770419  
; GENERAL INFORMATION:  
; APPLICANT: Xu, Feng  
; APPLICANT: Berka, Randy M.  
; APPLICANT: Wahlechner, Jill A.  
; TITLE OF INVENTION: BLUE COPPER OXIDASE MUTANTS WITH  
; TITLE OF INVENTION: ENHANCED ACTIVITY  
; NUMBER OF SEQUENCES: 29  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: No. 57704199 No. 5770419disk of No. 5770419th America, Inc.  
; STREET: 405 Lexington Avenue, 64th Floor  
; CITY: New York  
; STATE: New York  
; COUNTRY: United States of America  
; ZIP: 10174-6401  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/706.037  
; FILING DATE: 30-AUG-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/002,800  
; FILING DATE: 1-SEP-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Harrington, James J.  
; REGISTRATION NUMBER: 38,711  
; REFERENCE/DOCKET NUMBER: 4526.200-US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-867-0123  
; TELEFAX: 212-878-9655  
; INFORMATION FOR SEQ ID NO: 25:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 529 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-706-037-25

Query Match 5.4%; Score 141; DB 1; Length 529;

Best Local Similarity 20.4%; Pred. No. 0.00019;  
Matches 103; Conservative 60; Mismatches 216; Indels 126; Gaps 20;

QY 59 LPVIDAIVTAPEVPPVDRDHPAKVVK--METVEKVMRLADGVEYQFMTFGQVPGQM 116  
Db 8 LPLAAVST-----PAPAFAVRVYKFKDKVNVVAPDGFQSGISVNGLVGTL 54  
QY 117 IRVREGTIEVQPSNH---PDSKRPANVDH---AATGGCGGAASFTAP-----GHTS 164  
Db 55 ITANKGTLRLINVTNQLTDPSSMRATTIHHGLFQATTADEDDGPAFVTCQPIAONLSITY 114

QY 165 TFSFKALQPLVYVHCAPVPGNHIANMGY-LILVEP----KEGLPKYDKEYVYNGDF 219  
Db 115 EIPLRG-QGTWVYH---AHLASQYVDGLRGLPLVIDPNDPHKSRVDVDDASTVWLEDM 170  
QY 220 Y-----TKGKGEQG---LQPFMEKAIREDAEYVNRNGSVGALTGENALKAKGVEY 269  
Db 171 YHTPAFVLEKOMSTNTNALLSPVDSGLINGKRVY-----GGPVPVRSVINVXKRGKY 225  
QY 270 RLPVNGGPNLTSSPFVIGEIPDKVHFBEGKGNNHIQTLIPAGS-----AITEFKVD 324  
Db 226 RLRLVITASAIGSTFETI-----EGHSL--TVLEADSLIQLPLAVDSFOIY 268  
QY 325 VPQDYVLVDHAIFFAPNKALGLIKYGEENHEIYSHKOTDAVYLEPAGQALDQGE--- 381  
Db 269 AGORYSVIVEANOTANWIRAPMTVAGAGTNALDPTNVFAVLHGYGADNAPTTEQGS 328  
QY 382 -----AKTPAPANLQOIKAGKATYDS-----NCAACHQPGKGV 417  
Db 329 AIGTALVEENLHALINPGAPGASAPADVSLNLAIGRSTVDGILRFTFNRIKYEAP----- 383  
QY 418 PNAFPPLANSDYLNADHARAASIVANGLSGITVNGQYESVWPAIALSDQOIANVITYT 477  
Db 384 --SLPTL-----LKLANNASNDADFTNEHTIYLPHNKVIELNITGADHP 428  
QY 478 LNSFGKGGQLSADDAVAKAKTKPN 502  
Db 429 IHLHG-----VFIDVKSLSGTFN 447

## RESULT 13

US-09-005-397-25  
; Sequence 25, Application US/09005397  
; Patent No. 5972670  
; GENERAL INFORMATION:  
; APPLICANT: Xu, Feng  
; APPLICANT: Berka, Randy M.  
; APPLICANT: Wahlechner, Jill A.  
; TITLE OF INVENTION: BLUE COPPER OXIDASE MUTANTS WITH  
; TITLE OF INVENTION: ENHANCED ACTIVITY  
; NUMBER OF SEQUENCES: 29  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: No. 59726700 No. 5972670disk of No. 5972670th America, Inc.  
; STREET: 405 Lexington Avenue, 64th Floor  
; CITY: New York  
; STATE: New York  
; COUNTRY: United States of America  
; ZIP: 10174-6401  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/005,397  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/706,037  
; FILING DATE: 30-AUG-1996  
; APPLICATION NUMBER: US 60/002,800  
; FILING DATE: 1-SEP-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Harrington, James J.  
; REGISTRATION NUMBER: 38,711  
; REFERENCE/DOCKET NUMBER: 4526.200-US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-867-0123  
; TELEFAX: 212-878-9655  
; INFORMATION FOR SEQ ID NO: 25:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 529 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear



```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/993,318A
FILING DATE: December 18, 1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Gregg, Valeta A.
REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 5032.200-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
TELEX:
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 529 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-993-318A-5

```

```

Query Match 5.4%; Score 141; DB 2; Length 529;
Best Local Similarity 20.4%; Pred. No. 0.00019;
Matches 103; Conservative 60; Mismatches 216; Indels 126; Gaps 20;

```

```

QY 59 LPVIDAIVTAPEVPPVDHDPKAVVK--METVEKWLADGVEQFTFGQVPGOM 116
DB 8 LPLLAAVST-----PAPAFAVNNYKFDIKNNVAPDPGFSISVNGLVPGTL 54
QY 117 IRVREGDTIEVQFSNH---PDSKMPHNVDPH---AATGGGGAASFTAP-----GHTS 164
DB 55 ITANKGDTLRINTVNTQLTDSMRKRAATTIMHGLFQATTADEDEGPAFTQCPFIQNLSTY 114
QY 165 TFSFKALQPLGLVYHCAVAPVGMHIANMGV-LILVER---KEGLPKVDKEYVVMQGF 219
DB 115 EIPLRG-QTGTMWYH---AHLASQYVDGLRGPLVIYDPNDPHKSRVYDDASTVWLEDM 170
QY 220 Y-----TKGKGEQG---LQPFDMKAIREDAEVYVNGSVALLGENNAKAKVGETV 269
DB 171 YHTPAPVLEKQMFSTNTTALLSPVDSGLINGKGRVY-----GGPAVPRSVINVKRGKRY 225
QY 270 RLFVNGGPNLTSSFFHVGIEIFDKVHEGKGNNHIOITLLIPAG-----AATFEKVD 324
DB 226 RLKRVINSAIGSFSTI-----EGHSL--TVIEDGILHQPLAVDSFQIY 268
QY 325 VPQGDYLVDAIFRAFNKAGALGILKEGEGNHEIYSHKQTDVAVYLPFGAPQALDTQGE 381
DB 269 AGQRYSVIVEANQTAANYWIRAPMTVAGAGTNANLDPNVFAVLHYEGAPNAEPTTEQGS 328
QY 382 -----AKTPAPANLQEQIKAGKATYDS-----NCAACHQPDGKGV 417
DB 329 AIGTALVEENLHALINPGAPGSAADVSLNLAIGSTVDGILRTFNNIKYAP-----383
QY 418 PNAFPLANSDYLNADHARAASIYANGLSGITVNGQYESVMPAIALSDQOIANVITYT 477
DB 384 --SLPTL-----LKIANNASNDADPTPEHITVLPHNKVIELNITGGADHP 428
QY 478 LNSFGNKGQSLADDAVAKAKTKPN 502
DB 429 IHLHG-----VPDIVKSLGQTN 447

```

```

RESULT 16
US-09-399-886-5
Sequence 5, Application US/09399886
Patent No. 6140092
GENERAL INFORMATION:
APPLICANT: Pedersen, Anders
APPLICANT: Svendsen, Allan
APPLICANT: Schneider, Paule

```

```

APPLICANT: Raasmussen, Grethe
APPLICANT: Cherry, Joel
TITLE OF INVENTION: LACCASE MUTANTS
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSER: No. 61400920 No. 6140092disk of No. 61400922th America
STREET: 405 Lexington Avenue
CITY: New York
COUNTRY: USA
ZIP: 10174
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/399,886
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/993,318
FILING DATE: December 18, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Gregg, Valeta A.
REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 5032.200-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
TELEX:
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 529 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-399-886-5

```

```

Query Match 5.4%; Score 141; DB 3; Length 529;
Best Local Similarity 20.4%; Pred. No. 0.00019;
Matches 103; Conservative 60; Mismatches 216; Indels 126; Gaps 20;

```

```

QY 59 LPVIDAIVTAPEVPPVDHDPKAVVK--METVEKWLADGVEQFTFGQVPGOM 116
DB 8 LPLLAAVST-----PAPAFAVNNYKFDIKNNVAPDPGFSISVNGLVPGTL 54
QY 117 IRVREGDTIEVQFSNH---PDSKMPHNVDPH---AATGGGGAASFTAP-----GHTS 164
DB 55 ITANKGDTLRINTVNTQLTDSMRKRAATTIMHGLFQATTADEDEGPAFTQCPFIQNLSTY 114
QY 165 TFSFKALQPLGLVYHCAVAPVGMHIANMGV-LILVER---KEGLPKVDKEYVVMQGF 219
DB 115 EIPLRG-QTGTMWYH---AHLASQYVDGLRGPLVIYDPNDPHKSRVYDDASTVWLEDM 170
QY 220 Y-----TKGKGEQG---LQPFDMKAIREDAEVYVNGSVALLGENNAKAKVGETV 269
DB 171 YHTPAPVLEKQMFSTNTTALLSPVDSGLINGKGRVY-----GGPAVPRSVINVKRGKRY 225
QY 270 RLFVNGGPNLTSSFFHVGIEIFDKVHEGKGNNHIOITLLIPAG-----AATFEKVD 324
DB 226 RLKRVINSAIGSFSTI-----EGHSL--TVIEDGILHQPLAVDSFQIY 268
QY 325 VPQGDYLVDAIFRAFNKAGALGILKEGEGNHEIYSHKQTDVAVYLPFGAPQALDTQGE 381
DB 269 AGQRYSVIVEANQTAANYWIRAPMTVAGAGTNANLDPNVFAVLHYEGAPNAEPTTEQGS 328
QY 382 -----AKTPAPANLQEQIKAGKATYDS-----NCAACHQPDGKGV 417
DB 329 AIGTALVEENLHALINPGAPGSAADVSLNLAIGSTVDGILRTFNNIKYAP-----383
QY 418 PNAFPLANSDYLNADHARAASIYANGLSGITVNGQYESVMPAIALSDQOIANVITYT 477

```

Db 384 --SLPTL-----LKLANNASNDADFTPNEHTTVLPHNKVIELNTGGADHP 428

QY 478 LNSFGKGGQLSADVDVAKAKTKPEN 502  
: : : : :  
Db 429 IHLHGH-----VFIVKSLGGTNP 447

RESULT 17  
US-09-396-260-5  
; Sequence 5, Application US/09396260  
; Patent No. 6184015  
; GENERAL INFORMATION:  
; APPLICANT: Svendsen, Allan  
; APPLICANT: Xu, Feng  
; TITLE OF INVENTION: LACCASE MUTANTS  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: No. 61840150 No. 6184015disk of No. 6184015th America, Inc.  
; STREET: 405 Lexington Avenue  
; CITY: New York  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10174  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSEO for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/396, 260  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/09/032,315  
; FILING DATE: 27-FEB-1998  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Rozek, Carol  
; REGISTRATION NUMBER: 36,993  
; REFERENCE/DOCKET NUMBER: 5200.200-US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-867-0123  
; TELEFAX: 212-878-9655  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; \*LENGTH: 529 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-09-396-260-5

Query Match 5.4%; Score 141; DB 3; Length 529;  
Best Local Similarity 20.4%; Pred. No. 0.00019;  
Matches 103; Conservative 60; Mismatches 216; Indels 126; Gaps 20;

QY 59 LPVIDAIVTHAPVPPVDRDHPAKVVK--METVEKVMKMLADGVEYQFTFGGQVPGQM 116  
: : : : :  
Db 8 LPLLAIVST-----PAPFAVNNYKFDIKVNVVAPDGFQSRISVNGLVPGTL 54

QY 117 IRVREGDTLEVQPSNH---PDSKMPHNVDPH---AATGPGGGAESFTAP-----GHTS 164  
: : : : :  
Db 55 ITANKGDTLRINTVNTQUTDPSMRATTIHMHGTFQATTADEDDPAFTQCPINQNSITY 114

QY 165 TFSFKALQPLGLYYHCAVAPVGHIANMGK-LILVER---KEGLPKVDKEYVVMQGF 219  
: : : : :  
Db 115 EIPLRG-QGTGMWYH---AHLASQYVDGLRGLPLVIYDNDPHKSRVDVDDASTVVMLEW 170

QY 220 Y-----TRKXVGEQG---LOPFDMKAIREDDAEYVFNFGVGLTGENALAKAVGETV 269  
: : : : :  
Db 171 YHTPAPVLEKOMSTNTTALLSPVDSGLINGKRYV---GGPAVPRSVINVKKGRY 225

QY 270 RLTVNGGPNLTSSFHVIGEIFPKVHEGGKGENHNIQTLLIPAGG-----AAITPKVD 324  
: : : : :  
Db 8 LPLLAIVST-----PAPFAVNNYKFDIKVNVVAPDGFQSRISVNGLVPGTL 54

Db 226 RLRVINASAIIGSTFESI-----EGHSL--TVIEADGILHOPLAVDSFOIY 268

QY 325 VPGDVLVDHALFPAFNKALGILKVEGENHEIYSHKOTDAVYLPREGAPQAIQOE--- 381  
: : : : :  
Db 269 AGGRYSVIVBANQNTANWYIRAPMTYAGAGTANLDPFTVFVLAHTEGAPNMEPTTEGSS 328

QY 382 -----APKTPAPANLQEOIKAGATYDS-----NCAACHQPDGKV 417  
: : : : :  
Db 329 AIGTALVEENLHALINPGAPGSGAPADVSINLAIGRSTVDGILRFTFNINIKYEAP----- 383

QY 418 PNAFPEPLANSDIYANAHAAASIVANGLSGKITVNGNOYESWPAIALSDQOQIANYITYT 477  
: : : : :  
Db 384 --SLPTL-----LKLANNASNDADFTPNEHTTVLPHNKVIELNTGGADHP 428

QY 478 LNSFGKGGQLSADVDVAKAKTKPEN 502  
: : : : :  
Db 429 IHLHGH-----VFIVKSLGGTNP 447

RESULT 18  
US-09-576-281-5  
; Sequence 5, Application US/09576281  
; Patent No. 6277611  
; GENERAL INFORMATION:  
; APPLICANT: Pedersen, Anders  
; APPLICANT: Svendsen, Allan  
; APPLICANT: Schneider, Palle  
; APPLICANT: Rasmussen, Gretie  
; APPLICANT: Cherry, Joel  
; TITLE OF INVENTION: LACCASE MUTANTS  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: No. 62776110 No. 6277611disk of No. 6277611th America  
; STREET: 405 Lexington Avenue  
; CITY: New York  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10174  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/576, 281  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/993,318  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Gregg, Valeta A.  
; REGISTRATION NUMBER: 33,728  
; REFERENCE/DOCKET NUMBER: 5032.200-US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-867-0123  
; TELEFAX: 212-878-9655  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; \*LENGTH: 529 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-09-576-281-5

Query Match 5.4%; Score 141; DB 3; Length 529;  
Best Local Similarity 20.4%; Pred. No. 0.00019;  
Matches 103; Conservative 60; Mismatches 216; Indels 126; Gaps 20;

QY 59 LPVIDAIVTHAPVPPVDRDHPAKVVK--METVEKVMKMLADGVEYQFTFGGQVPGQM 116  
: : : : :  
Db 8 LPLLAIVST-----PAPFAVNNYKFDIKVNVVAPDGFQSRISVNGLVPGTL 54







Db 338 GLLSHGGLKARMBEGSHHLSRSVVRKONETTIVVMDSEKLVLEYPGAACGSKPADLV 397  
Qy 393 EQIKAGKATYDSNCAACHQPDGKGVNAPPL-----ANSDYLNADHARAAS 439  
Db 398 LDL-----TFGLNFATGHWMI-NGIPYESPKIPTLLKILTDEDOGVTSDFTEEH---TV 448  
Qy 440 IVANGLSGKITVNGNOYESVMPALALSDQOIANVITYTLNSFCN 483  
Db 449 ILPKNKICIEFNKNGSGIPITHPVHLGH-----TWDVVOFCN 486

RESULT 23  
US-09-399-886-8  
Sequence 8, Application US/09399886  
Patent No. 6140092  
GENERAL INFORMATION:  
APPLICANT: Pedersen, Anders  
APPLICANT: Svendsen, Allan  
APPLICANT: Schneider, Paule  
APPLICANT: Rasmussen, Grethe  
APPLICANT: Cherry, Joel  
TITLE OF INVENTION: LACCASE MUTANTS  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 61400920 No. 6140092disk of No. 6140092ch America  
STREET: 405 Lexington Avenue  
CITY: New York  
COUNTRY: USA  
ZIP: 10174  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30 (ERO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/399,886  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/993,318  
FILING DATE: December 18, 1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Gregg, Valeta A.  
REGISTRATION NUMBER: 33,728  
REFERENCE/DOCKET NUMBER: 5032.200-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-867-0123  
TELEFAX: 212-878-9655  
TELEX:  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 575 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-399-886-8

Query Match 5.1%; Score 135; DB 3; Length 575;  
Best Local Similarity 20.0%; Pred. No. 0.00077;  
Matches 105; Conservative 63; Mismatches 176; Indels 180; Gaps 25;

Qy 86 VKMETVKKWRLA-----DSGEVQFTFGGQVGMIRREGDTIVQSN---HPDSK 136  
Db 17 VLARTVEYGLTISGEIAPDGVKNATLVNGVPGPLIPANKGTLVKVQNKLTNPEMY 76  
Qy 137 MPNVNDEFAA-----TFPGGAEASFT-----AFGHTSFSPKA-LOPGLVYVHCAV 182  
Db 77 RTTSHHGHGLQHNADDGP-----SFTVQCRIVRESESTYIPIPDQGTGYWH--- 127  
Qy 183 APVGMTHANGMYGLILVEPKK---GLPKVDKEYVVM-QGDFYTKGKGEGLOPFDMEKA 238  
Db 128 SHLSQVVDGLRGPLVIYKPKDPRRLVVDVDEKTVLITIGWY-----HESSKA 175

Qy 239 I-----REDAYVFNCSVGLTGEN-----ALKAKVGETVRLFVNGGPNLTS 282  
Db 176 ILASGNITQRPVSATINCK-GRFPDNTPPADPTLYTLKVKRGKRYRLRVINSSSIASF 234  
Qy 283 SFHVIGELFDKVFEGCKENHNIOITLLIPAGGAITEKRVV----- 325  
Db 235 RFSV-----EGHKV--TVIAAGVSTKPYQVDAFDILAQRIDCIVEANQ 277  
Qy 326 -PGDYVL-----VDHAFRAN-----KCAL 345  
Db 278 EPDTWINAPLNVNPKTAQALLVYEDRRPYHPKPYRKMSVSEALIKYMHKHKGR 337  
Qy 346 GILKVEG-----EENHEIYS-----HKQTDVAVLPGAPQALDTPQAPPTPAPANQ 392  
Db 338 GLLSHGGLKARMBEGSHHLSRSVVRKONETTIVVMDSEKLVLEYPGAACGSKPADLV 397  
Qy 393 EQIKAGKATYDSNCAACHQPDGKGVNAPPL-----ANSDYLNADHARAAS 439  
Db 398 LDL-----TFGLNFATGHWMI-NGIPYESPKIPTLLKILTDEDOGVTSDFTEEH---TV 448  
Qy 440 IVANGLSGKITVNGNOYESVMPALALSDQOIANVITYTLNSFCN 483  
Db 449 ILPKNKICIEFNKNGSGIPITHPVHLGH-----TWDVVOFCN 486

RESULT 24  
US-09-396-260-8  
Sequence 8, Application US/09396260  
Patent No. 6184015  
GENERAL INFORMATION:  
APPLICANT: Svendsen, Allan  
APPLICANT: Xu, Peng  
TITLE OF INVENTION: LACCASE MUTANTS  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 61840150 No. 6184015disk of No. 6184015ch America, Inc.  
STREET: 405 Lexington Avenue  
CITY: New York  
STATE: NY  
COUNTRY: USA  
ZIP: 10174  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/396,260  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/09/032,315  
FILING DATE: 27-FEB-1998  
ATTORNEY/AGENT INFORMATION:  
NAME: Rozek, Carol  
REGISTRATION NUMBER: 36,993  
REFERENCE/DOCKET NUMBER: 5200.200-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-867-0123  
TELEFAX: 212-878-9655  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 575 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-396-260-8

Query Match 5.1%; Score 135; DB 3; Length 575;  
Best Local Similarity 20.0%; Pred. No. 0.00077;  
Matches 105; Conservative 63; Mismatches 176; Indels 180; Gaps 25;







LENGTH: 540 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-687-580B-7

Query Match 4.9%; Score 129; DB 3; Length 540;  
 Best Local Similarity 22.3%; Pred. No. 0.0025;  
 Matches 100; Conservative 47; Mismatches 198; Indels 104; Gaps 20;

QY 100 GVEVQFTFGGVGGMIRREGDTIEVQSNHSDSPHNVDHATGEGGAEASF-- 157  
 DB 42 GVPOQGLINGQFPGPDYISVTNLLINFNSLDE--PILLSVNGIQRNRSFEDVYG 99  
 QY 158 ----TAGHTSTSFKAL-QPGLVYHCAVAPVGMHANGMYG--LIVEPKGLPRVD- 209  
 DB 100 TTCEIPFGKRPFTYLLQKDDQSGYYFPSLA---FHQAAGFGGKILSRPILVPPDP 156  
 QY 210 -KEYVMOGDFYTKGKGEGGL-----QPFDEKAIREDAEVYVNGSVGALTGENALK 262  
 DB 157 ADYTLTLIGDMYKKNHTALKAILDGGKKLPF-----PDGILLNG---RGPNGVS 202  
 QY 263 AKV--GETVALFVNGCPNLTSSPHVIGELFDKVFEGGKGNHTITLIPAGGAITE 320  
 DB 203 FVEQGGKTYRLRISNVLQNSLNRIGHRMKLVEVG---THLTQTYSSLDVHVGOT 258  
 QY 321 FKVDVPGDYVLVDHAI-----FRAFNKAGALILKEGEEHNEYSHKQTDVAVLPEGAP- 374  
 DB 259 YSVLITDQRAKHIVYVSSRFTSQVLTITGVHL-----YSNNTFVSGPFPBGPT 309  
 QY 375 -----QALDTQAPKTPAD-----ANLQEOIKA-----GKATYDNC 406  
 DB 310 IQIDWSLNQARSITRNLSASGPRENPOGSHYGMINTTRTIRLASAGQVNGKORAYVNS 369  
 QY 407 AACHQPGKGVNPAFPPLANSDYLNADHABAASL-----VANGLSGKITVNGQYESMPA 462  
 DB 370 VSEVPILD-----TPKLKLDYFKVGGFVSGISDAPSGGGLFLDTSVLGADYRQET 421  
 QY 463 IALSDQOIANYITYTLNSF-----GNKKG 486  
 DB 422 VFENTEDI--VQSMHLNGYGFVWVGMDGG 448

## RESULT 31

US-09-252-991A-23088  
 ; Sequence 23088, Application US/09252991A  
 ; Patent No. 6551795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ; FILE REFERENCE: 107196.136  
 ; CURRENT APPLICATION NUMBER: US/09/252.991A  
 ; CURRENT FILING DATE: 1999-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/074,788  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/094,190  
 ; PRIOR FILING DATE: 1998-07-27  
 ; NUMBER OF SEQ. ID NOS: 33142  
 ; SEQ. ID NO 23088  
 ; LENGTH: 310  
 ; TYPE: PRT  
 ; ORGANISM: Pseudomonas aeruginosa  
 US-09-252-991A-23088

Query Match 4.8%; Score 125.5; DB 4; Length 310;  
 Best Local Similarity 22.0%; Pred. No. 0.0022;  
 Matches 68; Conservative 37; Mismatches 89; Indels 115; Gaps 15;

QY 187 MHIANMGYGLIVPEKRGKLPKVDKEYVVMGDFYTKRGYGGGLQPFDEKAIREDAEV 246  
 DB 105 MNIAQEKYGPFA-----KYAMPLEEVAKDEHA-----MKQSGMFTATYC 145

QY 247 -VENG--VGALTGENALAKVGETVRLFVNGCPNLTSSPHVIGELFDKVFEGGKGEN 303  
 DB 146 SICHSGDAGCAL-----GFPULADN-----EKRWG-GDP 173  
 QY 304 HNIQTLLIPAGGAITEFKVDVPGDYVLVDHAIFFAPFNKAGALILKEGEEHNEYSHKQ 363  
 DB 174 QSIETITL--GGR-----HALMAWG-----DILGEDGVKNVAAVYRTE 210  
 QY 364 TDVAVLPEGAPQALDTQAPKTPAPANLQEOIKAGATYDNCACADPQDK-----GVY 418  
 DB 211 LAGLKLPBG-----KADVAGQITSVNCVACHGEBEGKTALVGP 252  
 QY 419 NAFPLANSDYLNADHARAASIVANGLSGKITVNGQYESVMPA--LSDOQIANVITY 476  
 DB 253 NLITNP--GAFYQSSVYAQLQOTIRHGRQO-----WPAQEPYLKKEKVTILAAV 259  
 QY 477 TLNFPNGK 485  
 DB 300 IYNLSHNG 308

## RESULT 32

US-09-032-315-7  
 ; Sequence 7, Application US/09032315  
 ; Patent No. 5985818  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Svendsen, Allan  
 ; TITLE OF INVENTION: LACCASE MUTANTS  
 ; NUMBER OF SEQUENCES: 10  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: No. 59858180 No. 5985818disk of No. 5985818th America, Inc.  
 ; STREET: 405 Lexington Avenue  
 ; CITY: New York  
 ; STATE: NY  
 ; COUNTRY: USA  
 ; ZIP: 10174  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: diskette  
 ; COMPUTER: IBM Compatible  
 ; OPERATING SYSTEM: DOS  
 ; SOFTWARE: FASTSEQ for Windows Version 2.0  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/032.315  
 ; FILING DATE: 27-FEB-1998  
 ; CLASSIFICATION:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Rozek, Carol  
 ; REGISTRATION NUMBER: 36,993  
 ; REFERENCE/DOCKET NUMBER: 5200.200-US  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 212-867-0123  
 ; TELEFAX: 212-878-9655  
 ; INFORMATION FOR SEQ. ID NO: 7:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 572 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 US-09-032-315-7

Query Match 4.7%; Score 124; DB 2; Length 572;  
 Best Local Similarity 21.4%; Pred. No. 0.0081;  
 Matches 79; Conservative 55; Mismatches 124; Indels 112; Gaps 20;

QY 86 VKAEYEVKWRLA-----DGEYQFTFGGVGGMIRREGDTIEVQFSN---HPDSK 136  
 DB 17 VLARTVEYNLKIISNGKIADGVBRDATLVNGGYPGLIPANKGDTLKVQKQLTNPDMY 76  
 QY 137 MPMNVDFHAA-----TGGGGAASFTAGHTSTSFK-ALQGLIVYHCAVAPVGM 187  
 DB 77 RTTSHMHGLDHRNADDDGPAFVVTOCPV-VPQASVYTMPPLDDQGTGYWY---SHLS 132

QY 188 HIANMGYGLIIVEKEGLPKV---DKEYYVMOGDFY-TRKG---YEGQLQPFDMK 237  
DB 133 QYVDGLRGLVLYDPKDPHRLDYDIDDEKTVLIIIDGMYHTSSKAILATNITLQOPDS-- 190  
QY 238 AIREDAEYVFNCSVALTGEN-----ALKAVGETVRLFVNGGPNLTSSPHVIG 288  
DB 191 -----ATINGK-GRPDNDTPANPNTLYTLKVRGKRYLRVINS--SALASL----- 235  
QY 289 EIFDKVHEGKGENHNITOTLLIPAGGAITEFKVDVPGDYLVDHAFRAFNKALGIL 348  
DB 236 ----RMSIGHK-----MTVIAADVSTKPYQVD----- 260  
QY 349 KVEGSENEIYSHKQTDVAYLPEGAPOAIDTOEAPKT---PAP-ANLQEQIRAKGATYDS 404  
DB 261 -----SFDILAGORIDAV-----VEANOEPDTYWINAPLNTVANKTAQALLIYED 305  
QY 405 NCACHQPDG 414  
DB 306 DRPPYHPKG 315

## RESULT 33

US-08-993-318A-7  
Sequence 7, Application US/08993318A  
Patent No. 5998353  
GENERAL INFORMATION:  
APPLICANT: Pedersen, Anders  
APPLICANT: Svendsen, Allan  
APPLICANT: Schneider, Palle  
APPLICANT: Rasmussen, Grethe  
APPLICANT: Cherry, Joel  
TITLE OF INVENTION: LACCASE MUTANTS  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 59983530 No. 5998353disk of No. 5998353ch America  
STREET: 405 Lexington Avenue  
CITY: New York  
COUNTRY: USA  
ZIP: 10174  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/993,318A  
FILING DATE: December 18, 1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Gregg, Valeta A.  
REGISTRATION NUMBER: 33,728  
REFERENCE/DOCKET NUMBER: 5032.200-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-867-0123  
TELEFAX: 212-878-9655  
TELEX:  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 572 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-993-318A-7

Query Match 4.7%; Score 124; DB 2; Length 572;

Best Local Similarity 21.4%; Pred. No. 0.0081;

Matches 79; Conservative 55; Mismatches 124; Indels 112; Gaps 20;

QY 86 VKMETVEKMRLLA-----DGVYQFWTGGGVPGOMIRVREDTILEVQFSN--HPDSK 136  
DB 17 VIATVEYMLKISNGKIAPDVERDATLVNGVYGPBLIFANKDITLKVKVQNTLTNPDMY 76

QY 137 MEHNVDFHAA-----TGPGGAESFTAGHTSTSEFK-ALQPLVYHCAVAPGM 187  
DB 77 RTTSHHMGHLLQHRNADDGPAFVTVQCPV-VPQASVYTMPLGDGTGYWH---SHLSS 132  
QY 188 HIANMGYGLIIVEKEGLPKV---DKEYYVMOGDFY-TRKG---YEGQLQPFDMK 237  
DB 133 QYVDGLRGLVLYDPKDPHRLDYDIDDEKTVLIIIDGMYHTSSKAILATNITLQOPDS-- 190  
QY 238 AIREDAEYVFNCSVALTGEN-----ALKAVGETVRLFVNGGPNLTSSPHVIG 288  
DB 191 -----ATINGK-GRPDNDTPANPNTLYTLKVRGKRYLRVINS--SALASL----- 235  
QY 289 EIFDKVHEGKGENHNITOTLLIPAGGAITEFKVDVPGDYLVDHAFRAFNKALGIL 348  
DB 236 ----RMSIGHK-----MTVIAADVSTKPYQVD----- 260  
QY 349 KVEGSENEIYSHKQTDVAYLPEGAPOAIDTOEAPKT---PAP-ANLQEQIRAKGATYDS 404  
DB 261 -----SFDILAGORIDAV-----VEANOEPDTYWINAPLNTVANKTAQALLIYED 305  
QY 405 NCACHQPDG 414  
DB 306 DRPPYHPKG 315

## RESULT 34

US-09-399-886-7  
Sequence 7, Application US/09399886  
Patent No. 6140092  
GENERAL INFORMATION:  
APPLICANT: Pedersen, Anders  
APPLICANT: Svendsen, Allan  
APPLICANT: Schneider, Palle  
APPLICANT: Rasmussen, Grethe  
APPLICANT: Cherry, Joel  
TITLE OF INVENTION: LACCASE MUTANTS  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 61400920 No. 6140092disk of No. 6140092ch America  
STREET: 405 Lexington Avenue  
CITY: New York  
COUNTRY: USA  
ZIP: 10174  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/399,886  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/993,318  
FILING DATE: December 18, 1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Gregg, Valeta A.  
REGISTRATION NUMBER: 33,728  
REFERENCE/DOCKET NUMBER: 5032.200-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-867-0123  
TELEFAX: 212-878-9655  
TELEX:  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 572 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-399-886-7





```

;
; TELEFAX: 212-878-9655
;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 572 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
;
; JS-09-576-281-7

```

Query Match	4.7%	Score 124;	DB 3;	Length 572;
Best local Similarity	21.4%	Pred NC	0.0081;	

```

OY      86 VKMEIVEXKMRLA-----DGYEYQWTEGGGVPGCMIRVRBGDTIEVQESN---HPDSK 136
      ||| ||| : : : : : ||| ||| : : : : : ||| ||| : : : : :
Db      17 VLARTVEYNLKTSNGKIADPQVERDATALVNGVYEGPLIFANKQDITLKVKQNTLTNPDMY 76
      ||| ||| : : : : : ||| ||| : : : : : ||| ||| : : : : :
OY      137 MPHNVDFFHAA-----TGPGGAGASFLAPHTSTESFK-ALQJGLYYHCAVAPVCM 187
      : : : ||| ||| : : : : : ||| ||| : : : : : ||| ||| : : : : :
Db      77 RTTSHHMGLOLRNADDDGPAFVTOCP1-VPOASTYTHPCLGDQGTGYTH---SHLSS 122
      : : : ||| ||| : : : : : ||| ||| : : : : : ||| ||| : : : : :
OY      188 HIANGMYG-LILVEPEGLPKV---DKEYVWQSDPY-TKGK---YGEQGLQPPDMEK 237
      ||| ||| : : : : : ||| ||| : : : : : ||| ||| : : : : :
Db      133 QYVDLQRPGLVLYDPKDPHRLYDIDDEKTVLIGDWMYHSSKAILATGNITLQOPDS-- 190
      ||| ||| : : : : : ||| ||| : : : : : ||| ||| : : : : :
OY      238 AIREDAEYVVPFGSGALTGEN-----ALKAVGETVLFVNGGPNLTSFHVIG 288
      ||| ||| : : : : : ||| ||| : : : : : ||| ||| : : : : :
Db      191 -----ATTNGK-GRPDNPPTPANPNLTLYTLKVRGKRRLRVINS--SATA5F--- 235
      ||| ||| : : : : : ||| ||| : : : : : ||| ||| : : : : :
OY      289 EIFDKVHEEGSGKGNHNHIDTTLIPAGGAITEPKVAVPGDYVLVDNAIFPAFNKALGIL 348
      : : : ||| ||| : : : : : ||| ||| : : : : : ||| ||| : : : : :
Db      236 -----RMSIQGHR-----MTYIADGVSTKPYQVD----- 260
      ||| ||| : : : : : ||| ||| : : : : : ||| ||| : : : : :
OY      349 KVEGSEENHEIYSHKOTDAVYLLEPGADQALDIOEAPKT---PAP-ANLQEOIKAGKATYDS 404
      : : : ||| ||| : : : : : ||| ||| : : : : : ||| ||| : : : : :
Db      261 -----SFIDILAQGRIDAV-----VEANGPEPTYMINAELTVNANTQAALLIYED 305
      : : : ||| ||| : : : : : ||| ||| : : : : : ||| ||| : : : : :
OY      405 NCAACHQPDG 414
      ||| ||| : : : : : ||| ||| : : : : : ||| ||| : : : : :
Db      306 DRRPYHPKQ 315
      ||| ||| : : : : : ||| ||| : : : : : ||| ||| : : : : :

```

RESULT 37  
 US-09-032-315-1  
 Sequence 1, Application US/09032315  
 Patent No. 5985818  
 GENERAL INFORMATION:  
 APPLICANT: Svendsen, Allan  
 APPLICANT: Xu, Feng  
 TITLE OF INVENTION: LACCASE MUTANTS  
 NUMBER OF SEQUENCES: 10  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: No. 5985818o No. 5985818dsk of No. 5985818th America, Inc.  
 STREET: 405 Lexington Avenue  
 CITY: New York  
 STATE: NY  
 COUNTRY: USA  
 ZIP: 10174  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette  
 COMPUTER: IBM Compatible  
 OPERATING SYSTEM: DOS  
 SOFTWARE: Fastseq for Windows Version 2.0  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/032,315  
 FILING DATE: 27-FEB-1998  
 CLASSIFICATION:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Rozek, Carol  
 REGISTRATION NUMBER: 36,993  
 REFERENCE/DOCKET NUMBER: 5200.200-US

```

? TELECOMMUNICATION INFORMATION
? TELEPHONE: 212-867-0123
?
? TELEFAX: 212-878-9655
?
? INFORMATION FOR SEQ ID NO: 1:
?   SEQUENCE CHARACTERISTICS:
?     LENGTH: 539 amino acids
?     TYPE: amino acid
?     STRANDEDNESS: single
?     TOPOLOGY: linear
?     MOLECULE TYPE: protein
US-09-032-315-1

```

Query Match	4.6%;	Score 121.5;	DB 2;	Length 539;
Best Local Similarity	23.6%;	Pred. No. 0.013;		
Matches	92;	Conservative	50;	Mismatches 165;
				Indels 83;
				Gaps 23;

```

OY 112 VPGOMIRREDDETTEVCSN---HDSKMPHNVDFHA-----ATGGCGAASFTAP 160
Db 49 VHGPILRGKRDNELVNDLDNPTMLRPSIHNGLEFGRTNMADGADGVNQCPI--SP 107
OY 161 GHTSTFSF-KALOPGLVYVHCVAVAPVGNHIANGYGLIV---EPKESLPVKDEYYM 215
Db 108 GHAFLYKPTPAGHAGTFWYH---SHFGYQCGLRGPMVIYDDNDPHALYDEDDENTII 164
OY 216 Q-GPEYTKGKGGEGLQPFDMEKALREDAEYVFNFGS---VCALIGE-NALKAKVE--T 268
Db 165 TLAWMYHLPASISG-----AAQPPA--TLINGKRYVGGPAAEELISVWVEQCKYR 214
OY 269 VRLPVGNGCPULTSSPHVIGELPDKVHEEGKGEGENHNIQTLLIPAGAAITEPKVUPVD 328
Db 215 MRLSLSCDDPWW---QFSIDGHELTITIEVDGQTEPHYDRLQIFGQARSFVLDAQP--- 270
OY 329 YVLVDHAIFFRAF-NKGAIGILKVEEENHEIYSHKQDAVVLPEGAPOAIDTQAEKTPA 387
Db 271 ---VDNWMYIRAQPKKGRNGLAGT-----FANGVNSALIKRYAGAAANDPTTSA--NPN 317
OY 388 PANTQEQIKAGKATYDSNCAACHQPDGCVPNAPFPPLANSBYLNLADHAPAASTIVANGLSG 447
Db 318 PAQLNE-----ADLHALIIPAAIGIPT--FGAAD---VNNRFLQFGFS-----GG 356
OY 448 KITVNGNGYESVMPAI-----ALSDQOLAN 472
Db 357 RFTINGTAYES--PSVPTLLQIMSGAQSAN 384

```

RESULT 38  
US-08-993-318A-1  
Sequence 1, Application US/08993318A  
Patent No. 5998353  
GENERAL INFORMATION:  
APPLICANT: Pedersen, Anders  
APPLICANT: Svendsen, Allan  
APPLICANT: Schneider, Palle  
APPLICANT: Rasmussen, Grethe  
APPLICANT: Cherry, Joel  
TITLE OF INVENTION: LACCASE MUTANTS  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 59983530 No. 5998353dlek of No. 5998353lh America  
STREET: 405 Lexington Avenue  
CITY: New York  
COUNTRY: USA  
ZIP: 10174  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/993\_318A  
FILING DATE: December 18, 1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:

NAME: Gregg, Valeta A.  
REGISTRATION NUMBER: 33,728  
REFERENCE/DOCKET NUMBER: 5032.200-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-867-0123  
TELEFAX: 212-878-9655  
TELEX:  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 539 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-993-318A-1

Query Match 4.6%; Score 121.5; DB 2; Length 539;  
Best Local Similarity 23.6%; Pred. No. 0.013;  
Matches 92; Conservative 50; Mismatches 165; Indels 83; Gaps 23;

QY 112 VPGQIRVREGDTIEVQFSN---HPDSKMPHNVDFHA-----ATGPGGAASFTAP 160  
DB 49 VHGPIIRGKNDNFELNVVNDLDPMTLRPTSIHMGFLRGTNWADGADVQCPI-SP 107  
QY 161 GHTSTFSF-KALQPLGVYVHCAPVGMHIANGMVGLIIV-----EPKGLPKYDKEYVM 215  
DB 108 GHAFLYKFTPAHAGTFWYH---SHFGTOYCDGLRGMVYVDNDPHALYDEDDENTII 164  
QY 216 Q-GDFTKKGKYGEOGLPFPMEKAIREDAYVVFNGS---VGALTGE-NALKAKYGE--T 268  
DB 165 TLADWVHIAPASIQ-----AAQPD--TLINGKRVYGGPAELSIIVNEOGKKYR 214  
QY 269 VRLFVNGGPNLTSSFFHVGIFDKYHFEKGKGNHNIOTLLI-PAGGAITEFKVDVPCD 328  
DB 215 MRLISLSCDNN--QPSIDHELTIIIVDQLTERTVDRLOIFGQRYSFVLDANQP-- 270  
QY 329 YVLVDHAIFFRAF-NKGALGILKVEGENHEIYSHKQTDVYVLEPGAQALDQEPKTPA 387  
DB 271 ---VDVWYIRAOQPNKGRNGLAGT-----FANGVNSAILRYAGANADPTTSA--NPN 317  
QY 388 PAULOQIKAGKATYDSNCAACHOPDGKGVNAPFPPLANSDYVINDHAPASIVANGSLG 447  
DB 318 PAQLNE-----ADLHALIDPAAPGIPF--PGAAD--VNLRFQGFSS-----CG 356  
QY 448 KITVNGOYESVMPAI-----ALSDQOIAN 472  
DB 357 RFTINGTAYES--PSVPTLIQIMSGAQSAN 384

RESULT 39  
US-08-689-421-27  
Sequence 27, Application US/08689421  
Patent No. 6008029  
GENERAL INFORMATION:  
APPLICANT: Yaver, Debbie S.  
APPLICANT: Brown, Kimberley M.  
APPLICANT: Kaupplinen, Sakari  
APPLICANT: Haklert, Torben P  
TITLE OF INVENTION: PURIFIED COPRINUS LACCASES AND NUCLEIC ACIDS  
TITLE OF INVENTION: ENCODING SAME  
NUMBER OF SEQUENCES: 36  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 60080290 No. 6008029disk of No. 6008029th America, Inc.  
STREET: 405 Lexington Avenue, 64th floor  
CITY: New York  
STATE: New York  
COUNTRY: United States of America  
ZIP: 10174-6401  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/689,421  
FILING DATE: 9-AUG-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Harrington, James J.  
REGISTRATION NUMBER: 38,711  
REFERENCE/DOCKET NUMBER: 4554.204-WO  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-867-0123  
TELEFAX: 212-878-9655  
INFORMATION FOR SEQ ID NO: 27:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 539 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-689-421-27

Query Match 4.6%; Score 121.5; DB 3; Length 539;  
Best Local Similarity 23.6%; Pred. No. 0.013;  
Matches 92; Conservative 50; Mismatches 165; Indels 83; Gaps 23;

QY 112 VPGQIRVREGDTIEVQFSN---HPDSKMPHNVDFHA-----ATGPGGAASFTAP 160  
DB 49 VHGPIIRGKNDNFELNVVNDLDPMTLRPTSIHMGFLRGTNWADGADVQCPI-SP 107  
QY 161 GHTSTFSF-KALQPLGVYVHCAPVGMHIANGMVGLIIV-----EPKGLPKYDKEYVM 215  
DB 108 GHAFLYKFTPAHAGTFWYH---SHFGTOYCDGLRGMVYVDNDPHALYDEDDENTII 164  
QY 216 Q-GDFTKKGKYGEOGLPFPMEKAIREDAYVVFNGS---VGALTGE-NALKAKYGE--T 268  
DB 165 TLADWVHIAPASIQ-----AAQPD--TLINGKRVYGGPAELSIIVNEOGKKYR 214  
QY 269 VRLFVNGGPNLTSSFFHVGIFDKYHFEKGKGNHNIOTLLI-PAGGAITEFKVDVPCD 328  
DB 215 MRLISLSCDNN--QPSIDHELTIIIVDQLTERTVDRLOIFGQRYSFVLDANQP-- 270  
QY 329 YVLVDHAIFFRAF-NKGALGILKVEGENHEIYSHKQTDVYVLEPGAQALDQEPKTPA 387  
DB 271 ---VDVWYIRAOQPNKGRNGLAGT-----FANGVNSAILRYAGANADPTTSA--NPN 317  
QY 388 PAULOQIKAGKATYDSNCAACHOPDGKGVNAPFPPLANSDYVINDHAPASIVANGSLG 447  
DB 318 PAQLNE-----ADLHALIDPAAPGIPF--PGAAD--VNLRFQGFSS-----CG 356  
QY 448 KITVNGOYESVMPAI-----ALSDQOIAN 472  
DB 357 RFTINGTAYES--PSVPTLIQIMSGAQSAN 384

RESULT 40  
US-09-399-886-1  
Sequence 1, Application US/09399886  
Patent No. 6140092  
GENERAL INFORMATION:  
APPLICANT: Pedersen, Anders  
APPLICANT: Svendsen, Allan  
APPLICANT: Schneider, Palle  
APPLICANT: Rasmussen, Grethe  
APPLICANT: Cherry, Joel  
TITLE OF INVENTION: LACCASE MUTANTS  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 61400920 No. 6140092disk of No. 6140092th America  
STREET: 405 Lexington Avenue  
CITY: New York  
COUNTRY: USA  
ZIP: 10174  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/399,886
; FILING DATE:
; CLASSIFICATION:
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: 08/993,318
; FILING DATE: December 18, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Gregg, Valera A.
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 5032.200-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 539 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-399-886-1

```

```

Query Match          4.6%; Score 121.5; DB 3; Length 539;
Best Local Similarity 23.6%; Pred. No. 0.013;
Matches 92; Conservative 50; Mismatches 165; Indels 83; Gaps 23;

```

```

QY 112 VEGQIRVREGDTIEVQFSN--HPDSKPHNVDFHA-----ATGGGGAASFTAP 160
   |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
DB 49 VHGPIRGCKNPNFELNVNVDLDPFMTLRPTSIHMHGLFQRTNWDAGDVNQCP1-SF 107
   |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
QY 161 GHTSTFSF-KALQPLIYVHCAPVGMHIANGMVGLIIV---EPKESLPKVDKEYVYM 215
   |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
DB 108 GHAPLYKFTPAAGAGTFWYH--SHFGTQYCDGLRGPWVIYDDNDPFAALYDEDDENTII 164
   |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
QY 216 Q-GDFYTKGKYGEQGLQPFMEKAIREDAYVVFNGS--VGALTGE-NALKAKVGE--T 268
   |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
DB 165 TLADYTHIPAPSIQ-----AAQPD--TLINGKRYVGGPAALSLIVNVEQKKYR 214
   |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
QY 269 VRLFVNGGPNLTSSPFHVGELFDKVFREGKGGENHIQTLLIPAGGAALTEFKVDVPGD 328
   |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
DB 215 MRLISLSCDPNW--QFSIDGHELTITIEVDGQLTEPHVDRLQIFTGORYSFVLDAWQP-- 270
   |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
QY 329 YVLVDHALFRAF-NKGALGILKVEGENHEIYSHKQTDVAVLPBGAPOAIDTQBAKTPA 387
   |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
DB 271 ---VDNYWIRAPQPNKGRNGLAGT-----FANGVNSAILRYAGANADPTTSA--NPN 317
   |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
QY 388 PANTQEQIKAGKATYDSNCAACHQPDGKGVNAPFPLANSDYLNADHARAASIVANGLSG 447
   |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
DB 318 PAQLNE-----ADLHALIDPAPAPPIPT--PGAAD--VNLRFQIGFS-----CG 356
   |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
QY 448 KITVNGNOYESVMPAI-----ALSDQOIAN 472
   |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
DB 357 PFTINGTAVES--PSVPTLLQIMSGAQSAN 384
   |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|

```

Search completed: August 27, 2003, 18:33:21  
Job time : 33 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 27, 2003, 18:19:41 ; Search time 84 Seconds

(without alignments)  
948,580 Million cell updates/sec

Title: US-10-088-045-2

Perfect score: 2628  
Sequence: 1 MSKPLIKITTLICALSALML.....NKGGQLSADVAKAKKTKEN 502

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

A.Geneseq.13Jun03.\*  
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*  
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2626	100.0	502	22	AA67668
2	2626	100.0	502	22	AA67669
3	1388	52.9	392	24	ABP77627
4	1034	39.4	197	23	AA017561
5	349	13.3	924	22	AA742578
6	237	9.0	133	21	AA742474
7	237	9.0	184	24	ABP78130
8	222	8.5	132	21	AA742476
9	222	8.5	133	21	AA742475

10	172	6.5	1209	22	AA48440
11	163	6.2	478	14	AA33935
12	163	6.2	478	21	AA33938
13	163	6.2	579	12	AA14306
14	162	6.2	478	14	AA33288
15	150	5.7	374	24	ABJ18767
16	148	5.6	478	20	AA14051
17	146.5	5.6	555	12	AA12339
18	146	5.6	467	13	AA20193
19	146	5.6	559	18	AA080880
20	145.5	5.5	468	12	AA13994
21	145	5.5	559	11	AA06518
22	144.5	5.5	581	23	AA076443
23	141	5.4	529	16	AA16301
24	141	5.4	529	18	AA16301
25	141	5.4	529	19	AA16310
26	141	5.4	529	19	AA60878
27	141	5.4	529	21	AA50728
28	140.5	5.4	493	22	AA693026
29	136.5	5.2	326	21	AA16365
30	136.5	5.2	557	19	AA43448
31	136.5	5.2	576	16	AA72318
32	135	5.1	575	19	AA62501
33	134	5.1	575	19	AA76313
34	132.5	5.0	541	23	AB91862
35	131.5	5.0	511	22	AA690799
36	131	5.0	497	22	AA679105
37	130	5.0	554	21	AA653112
38	130	5.0	554	23	AB92295
39	130	5.0	2183	21	AA803533
40	129.5	4.9	439	24	ABP78252
41	129.5	4.9	504	22	ABM00012
42	129	4.9	518	16	AA78680
43	129	4.9	540	16	AA78679
44	129	4.9	549	21	AA649530
45	129	4.9	585	20	AA105677

#### ALIGNMENTS

RESULT 1	AA67668	standard; Protein; 502 AA.
ID	AA67668	
AC	AA67668	
DT	11-JUN-2001	(first entry) :
DE	Amino acid sequence of a BASB109 polypeptide.	
KW	BASB109; bacterial infection; vaccine; genetic immunisation.	
XX	Moraxella catarrhalis.	
OS	Moraxella catarrhalis.	
PN	MO200119996-A1.	
PD	22-MAR-2001.	
PF	14-SEP-2000; 2000WO-EP09035.	
PR	14-SEP-1999; 99GB-0021691.	
PA	(SMK) SMITHKLINE BEECHAM BIOLOGICALS.	
PI	Thomard J.	
DR	WPI: 2001-257883/26.	
DR	N-PSDB; AAF55658.	
PT	Novel BASB109 polypeptides of Moraxella catarrhalis useful for	
PT	diagnostic, prophylactic and therapeutic purposes against microbial	
PT	diseases, preferably bacterial infections	

2-keto-D-glucuronate  
Cytocrome C. Gu  
Ascorbate oxidase  
Sequence of cytoch  
Pseudomonas aerugi  
G. oxydans D-601  
ADH complex protei  
Brassica napus mic  
A. altocetigenes me  
Brassica microspor  
Novel Botrytis cin  
Laccase RSLact. R  
Rhizoctonia solani  
Rhizoctonia solani  
Rhizoctonia solani  
R. solani Laccase  
C glutamicum prote  
Eucalyptus grandis  
Tobacco laccase c1  
Laccase RSLact. R  
Rhizoctonia solani  
Rhizoctonia solani  
Rhizoctonia solani  
Herbicidally activ  
Corynebacterium g1  
Corynebacterium g1  
Murtine factor V SE  
N. gonorrhoeae ant  
Coprinus cinereus  
Tobacco leaf antif  
Tobacco leaf antif  
Arbidopsis thalia  
Maize diphenyl oxi

XX Claim 3; Page 65; 93pp; English.

CC The present sequence represents a BASB109 polypeptide of Moraxella  
CC catarrhalis. BASB109 polypeptides and polynucleotides are useful for  
CC treating bacterial infections, and as research reagents and materials  
CC for the treatment of and diagnosis of diseases, particularly human  
CC diseases. They are useful for inducing an immune response in an  
CC individual, and to assess the binding of small molecule substrates and  
CC ligands in, for e.g. cells, cell-free preparations, chemical libraries,  
CC and natural product mixtures. BASB109 polynucleotides are useful for  
CC therapeutic or prophylactic purposes, in particular genetic immunisation  
CC and in diagnosis of the stage and type of infection.

XX Sequence 502 AA;

Query Match 100.0%; Score 2626; DB 22; Length 502;  
Best Local Similarity 100.0%; Pred. No. 2.9e-227;  
Matches 502; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSKPTLTKITLICALSALMLSGCSNQADKAQPKSSTVDAAKTANADNNAASQEHQGLP 60  
DB 1 MSKPTLTKITLICALSALMLSGCSNQADKAQPKSSTVDAAKTANADNNAASQEHQGLP 60  
QY 61 VIDAIYTHAPEVPPVDRDHPAKVVMETVEKVMRLADGVEYQFTFGQVPGQMRVR 120  
DB 61 VIDAIYTHAPEVPPVDRDHPAKVVMETVEKVMRLADGVEYQFTFGQVPGQMRVR 120  
QY 121 EGDTEVQFSNHPDSKMPHNVDFHAATGPGGAASFTAPGHTSTSFKALQPLGYVHC 180  
DB 121 EGDTEVQFSNHPDSKMPHNVDFHAATGPGGAASFTAPGHTSTSFKALQPLGYVHC 180  
QY 181 AVAPVGMHIANGWYGLILVEPKGLPKVDKEYVMQGDFTYKKGEGGLQPFDMKPAIR 240  
DB 181 AVAPVGMHIANGWYGLILVEPKGLPKVDKEYVMQGDFTYKKGEGGLQPFDMKPAIR 240  
QY 241 EDAEYVFNQSVGALTGEMALAKAVGETVRLFVNGGPNLTSFPHVIGELFDKVHFEQK 300  
DB 241 EDAEYVFNQSVGALTGEMALAKAVGETVRLFVNGGPNLTSFPHVIGELFDKVHFEQK 300  
QY 301 GENHNIOTTLIPAGGAITEFKVDVPGDYLVVDAIFRAFNKALGILKVEGSENEHIYS 360  
DB 301 GENHNIOTTLIPAGGAITEFKVDVPGDYLVVDAIFRAFNKALGILKVEGSENEHIYS 360  
QY 361 HKQTDVYVLEPGAPQALDTEAPKTPAPANLQEOIKAGKATYDSNCAACHQPDGKVPNA 420  
DB 361 HKQTDVYVLEPGAPQALDTEAPKTPAPANLQEOIKAGKATYDSNCAACHQPDGKVPNA 420  
QY 421 FPPPLANSDYLNADHARAASIVANGLSGKITVNGNQSVMPALALSDQOIANVITYTLNS 480  
DB 421 FPPPLANSDYLNADHARAASIVANGLSGKITVNGNQSVMPALALSDQOIANVITYTLNS 480  
QY 481 FGNGGQLSADDAVAKAKTKPN 502  
DB 481 FGNGGQLSADDAVAKAKTKPN 502

RESULT 2  
AAB67669  
ID AAB67669 standard; Protein; 502 AA.  
XX  
AC AAB67669;  
XX  
DT 11-JUN-2001 (first entry)  
XX  
DE Amino acid sequence of a BASB109 polypeptide.  
XX  
KM BASB109; bacterial infection; vaccine; genetic immunisation.  
XX  
OS Moraxella catarrhalis.  
XX  
PN WO200119996-A1.  
XX

PD 22-MAR-2001

XX 14-SEP-2000; 2000MO-EP09035.

XX 14-SEP-1999; 99GB-0021691.

XX (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.

XX Thonard J;

XX WPI; 2001-257883/26.

XX N-PSDB; AAF55657.

PT Novel BASB109 polypeptides of Moraxella catarrhalis useful for

PT diagnostic, prophylactic and therapeutic purposes against microbial

PS diseases, preferably bacterial infections

XX Claim 3; Page 66; 93pp; English.

CC The present sequence represents a BASB109 polypeptide of Moraxella  
CC catarrhalis. BASB109 polypeptides and polynucleotides are useful for  
CC treating bacterial infections, and as research reagents and materials  
CC for the treatment of and diagnosis of diseases, particularly human  
CC diseases. They are useful for inducing an immune response in an  
CC individual, and to assess the binding of small molecule substrates and  
CC ligands in, for e.g. cells, cell-free preparations, chemical libraries,  
CC and natural product mixtures. BASB109 polynucleotides are useful for  
CC therapeutic or prophylactic purposes, in particular genetic immunisation  
CC and in diagnosis of the stage and type of infection.

XX Sequence 502 AA;

Query Match 100.0%; Score 2626; DB 22; Length 502;  
Best Local Similarity 100.0%; Pred. No. 2.9e-227;  
Matches 502; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSKPTLTKITLICALSALMLSGCSNQADKAQPKSSTVDAAKTANADNNAASQEHQGLP 60  
DB 1 MSKPTLTKITLICALSALMLSGCSNQADKAQPKSSTVDAAKTANADNNAASQEHQGLP 60  
QY 61 VIDAIYTHAPEVPPVDRDHPAKVVMETVEKVMRLADGVEYQFTFGQVPGQMRVR 120  
DB 61 VIDAIYTHAPEVPPVDRDHPAKVVMETVEKVMRLADGVEYQFTFGQVPGQMRVR 120  
QY 121 EGDTEVQFSNHPDSKMPHNVDFHAATGPGGAASFTAPGHTSTSFKALQPLGYVHC 180  
DB 121 EGDTEVQFSNHPDSKMPHNVDFHAATGPGGAASFTAPGHTSTSFKALQPLGYVHC 180  
QY 181 AVAPVGMHIANGWYGLILVEPKGLPKVDKEYVMQGDFTYKKGEGGLQPFDMKPAIR 240  
DB 181 AVAPVGMHIANGWYGLILVEPKGLPKVDKEYVMQGDFTYKKGEGGLQPFDMKPAIR 240  
QY 241 EDAEYVFNQSVGALTGEMALAKAVGETVRLFVNGGPNLTSFPHVIGELFDKVHFEQK 300  
DB 241 EDAEYVFNQSVGALTGEMALAKAVGETVRLFVNGGPNLTSFPHVIGELFDKVHFEQK 300  
QY 301 GENHNIOTTLIPAGGAITEFKVDVPGDYLVVDAIFRAFNKALGILKVEGSENEHIYS 360  
DB 301 GENHNIOTTLIPAGGAITEFKVDVPGDYLVVDAIFRAFNKALGILKVEGSENEHIYS 360  
QY 361 HKQTDVYVLEPGAPQALDTEAPKTPAPANLQEOIKAGKATYDSNCAACHQPDGKVPNA 420  
DB 361 HKQTDVYVLEPGAPQALDTEAPKTPAPANLQEOIKAGKATYDSNCAACHQPDGKVPNA 420  
QY 421 FPPPLANSDYLNADHARAASIVANGLSGKITVNGNQSVMPALALSDQOIANVITYTLNS 480  
DB 421 FPPPLANSDYLNADHARAASIVANGLSGKITVNGNQSVMPALALSDQOIANVITYTLNS 480  
QY 481 FGNGGQLSADDAVAKAKTKPN 502  
DB 481 FGNGGQLSADDAVAKAKTKPN 502



Sequence 197 AA;  
 Query Match 39.4%; Score 1034; DB 23; Length 197;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-84;  
 Matches 197; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSKRTLLIKTLLICALSALMLSGCSNQADKAAQPKSSTVDAATAKANDNNAASQEHQELP 60  
 DB 1 MSKRTLLIKTLLICALSALMLSGCSNQADKAAQPKSSTVDAATAKANDNNAASQEHQELP 60  
 QY 61 VIDAIYTHAPBPVPVPRDHPAKVYKVEYKEMRLADGVYQFMTFGQVPGQMIRVR 120  
 DB 61 VIDAIYTHAPBPVPVPRDHPAKVYKVEYKEMRLADGVYQFMTFGQVPGQMIRVR 120  
 QY 121 EGDTEVQFSNHPDSKPHNVDFHAATGPGGAEASTAPGHTSTFSFKALQPLYYVHC 180  
 DB 121 EGDTEVQFSNHPDSKPHNVDFHAATGPGGAEASTAPGHTSTFSFKALQPLYYVHC 180  
 QY 181 AVAPVGMHIANGYGLI 197  
 DB 181 AVAPVGMHIANGYGLI 197

RESULT 5  
 AAU42578 standard; Protein; 924 AA.  
 ID AAU42578  
 AC AAU42578;  
 XX 27-FEB-2002 (first entry)  
 DT  
 DE Propionibacterium acnes immunogenic protein #3474.  
 XX  
 KM SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;  
 KM uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
 KM inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
 KM dermatological; osteopathic; neuroprotectant.  
 XX  
 OS Propionibacterium acnes.  
 XX  
 PN WO200181581-A2.  
 PD 01-NOV-2001.  
 PF 20-APR-2001; 2001WO-US12865.  
 PR 21-APR-2000; 2000US-199047P.  
 PR 02-JUN-2000; 2000US-208841P.  
 PR 07-JUL-2000; 2000US-216747P.  
 XX  
 PA (CORI-) CORIXA CORP.  
 XX  
 PI Sheikh YM, Persing DH, Mitcham JL, Wang SS, Bharia A;  
 PI L'Almeida J, Zhang Y, Jen S, Carter D;  
 DR N-PSDB; AAS59518.  
 DR WPI; 2001-616774/71.  
 XX  
 PT Propionibacterium acnes polypeptides and nucleic acids useful for  
 PT vaccinating against and diagnosing infections, especially useful for  
 PT treating acne vulgaris -  
 XX  
 PS Claim 3; SEQ ID NO 3773; 1069pp; English.  
 CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic  
 CC polypeptides. The proteins and their associated DNA sequences are used in  
 CC the treatment, prevention and diagnosis of medical conditions caused by  
 CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,  
 CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.  
 CC P. acnes is also involved in infections of bone, joints and the central  
 CC nervous system, however it is particularly involved in the inflammatory  
 CC lesions associated with acne vulgaris. A method for detecting the  
 CC presence or absence of P. acnes in a patient comprises contacting a

CC sample with a binding agent that binds to the proteins of the invention  
 CC and determining the amount of bound protein in the sample. The  
 CC polypeptides may be used as antigens in the production of antibodies  
 CC specific for P. acnes proteins. These antibodies can be used to  
 CC downregulate expression and activity of P. acnes polypeptides and  
 CC therefore treat P. acnes infections. The antibodies may also be used as  
 CC diagnostic agents for determining P. acnes presence, for example, by  
 CC enzyme linked immunosorbent assay (ELISA).  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at fdp.wipo.int/pub/published\_pct\_sequences.

Sequence 924 AA;  
 Query Match 13.3%; Score 349; DB 22; Length 924;  
 Best Local Similarity 31.4%; Pred. No. 5e-22;  
 Matches 92; Conservative 46; Mismatches 121; Indels 34; Gaps 10;

QY 72 VPPPV--DRDHPAKVYKVEYKEMRLADGVYQFMTFGQVPGQMIRVREGDTIEVQF 129  
 DB 650 VLPPLMTGRVHRMTLIAQ---ESVQELAPETITDAMTYNGRYMAPIHARIGDEMVRHL 705  
 QY 130 SNHPDSKPHNVDFHAATGPGGAEAST-----APGHTSTFSFKALQPLYYVHC 183  
 DB 706 VNR--GTMGSLDFHAGT-----VSPTRVMETIAPGQELDYNFTLHRAIGLWYHCSTA 756  
 QY 184 PVGMHIANGYGLIYKVEYKEMRLADGVYQFMTFGQVPGQMIRVREGDTIEVQF 243  
 DB 757 PMSAHIAAGFGAVIYPPHD-LPRADREFLVQSEYTLSEHNAE----VNTAKINETP 811  
 QY 244 EYVFNFGSVALTGENALPAKVGETVRLFPVNGGPNLTSSFVHIGEIFDKVHEG----- 298  
 DB 812 DLTMFNGHANQVYFE-FLKARVERVRIWLAAGPBGCSFHVVGQFDTVFEAGAYTLK 870  
 QY 299 -GKGENHNTQTILIPAGGAITEFKVDVPGDYLVVHAIFRANKALGLKY 350  
 DB 871 RGNPEGGCGCALDLASAGGFVEMVEBERGYTFVMS-FVEMERGAKEFIEV 922

RESULT 6  
 AAU74274 standard; Protein; 133 AA.  
 ID AAU74274  
 AC AAU74274;  
 XX 21-MAR-2000 (first entry)  
 DT  
 DE Neisseria gonorrhoeae ORF 007 protein sequence SEQ ID NO:44.  
 XX  
 KM Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;  
 KM antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;  
 KM antibacterial; gene therapy.  
 XX  
 OS Neisseria gonorrhoeae.  
 XX  
 PN WO9957280-A2.  
 PD 11-NOV-1999.  
 PF 30-APR-1999; 99WO-US09346.  
 PR 01-MAY-1998; 98US-0083758.  
 PR 31-JUL-1998; 98US-0094869.  
 PR 02-SEP-1998; 98US-0098994.  
 PR 02-SEP-1998; 98US-0099062.  
 PR 09-OCT-1998; 98US-0103749.  
 PR 09-OCT-1998; 98US-0103794.  
 PR 09-OCT-1998; 98US-0103796.  
 PR 25-FEB-1999; 99US-0121528.  
 XX  
 PA (CHIR) CHIRON CORP.  
 PA (GENO-) INST GENOMIC RES.





CC represent novel *Neisseria meningitidis* and *N. gonorrhoeae* polynucleotides  
CC and polypeptides. AA254537 to AA254576 and AA254616 to AA25473 represent  
CC PCR primers used in the exemplification of the present invention. The  
CC polypeptides, the polynucleotides, antibodies and compositions of  
CC the invention can be used as vaccines, as diagnostic reagents, and as  
CC immunogenic compositions. The polypeptides can be used in the  
CC manufacture of medicaments for treating or preventing infection due to  
CC *Neisseria* bacteria (e.g. meningitis and septicemia), to detect the  
CC presence of *Neisseria* bacteria, or to raise antibodies. They may also  
CC be used to screen for agonists or antagonists, which may themselves  
CC have use as antibacterial agents. The polynucleotides of the invention  
CC may also be used in gene therapy protocols.

XX Sequence 132 AA;

Query Match 8.5%; Score 222; DB 21; Length 132;

Best Local Similarity 42.7%; Pred. No. 7.3e-12;

Matches 44; Conservative 14; Mismatches 45; Indels 0; Gaps 0;

Db 398 GKATYDSCAACHQPDGKGVNAPPLANSPLYLNADHARAASIVANGSGKITVNGNOYE 457

Qy 30 GQKYYESNCVACHGKKGEGRTMFPPLYSDFIMKKQVLLSHVKGINGTIVNGKTYN 89

Db 458 SVMPAIALSDQOIANVITTYLNSFGNKGQLSADVDVAKAKKTK 500

Qy 90 GFWPATIAISDADIAAVALTYIMNAPDNGGSGVTEKDVQAKKKK 132

RESULT 9

AAV74275  
ID AAV74275 standard; Protein; 133 AA.

XX AAV74275;

XX 21-MAR-2000 (first entry)

DE *Neisseria meningitidis* ORF 007 protein sequence SEQ ID NO:46.

XX *Neisseria meningitidis*; *Neisseria gonorrhoeae*; antigen; vaccine;

KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;

XX antibacterial; gene therapy.

XX *Neisseria meningitidis*.

OS *Neisseria meningitidis*.

XX WO957280-A2.

XX 11-NOV-1999.

XX 30-APR-1999; 99WO-US09346.

XX 01-MAY-1998; 98US-0083758.

XX 31-JUL-1998; 98US-0094869.

XX 02-SEP-1998; 98US-0098994.

XX 02-SEP-1998; 98US-0099062.

XX 09-OCT-1998; 98US-0103749.

XX 09-OCT-1998; 98US-0103794.

XX 09-OCT-1998; 98US-0103796.

XX 25-FEB-1999; 99US-0121528.

XX (CHIR) CHIRON CORP.

XX (GENO-) INST GENOMIC RES.

XX Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M,

XX Pieteren J, Piza M, Rappuoli R, Ratti G, Scalato E, Scarelli M,

XX Tetrein H, Venter JC;

XX WPI; 2000-062150/05.

XX N-PSDB; AA253037.

XX Novel *Neisseria* polypeptides predicted to be useful antigens for

XX PT vaccines and diagnostics

XX Claim 2; Page 180; 1453P; English.

XX AA253015 to AA254536, AA254577 to AA254615, and AAV74253 to AAV75941  
CC represent novel *Neisseria meningitidis* and *N. gonorrhoeae* polynucleotides  
CC and polypeptides. AA254537 to AA254576 and AA254616 to AA25473 represent  
CC PCR primers used in the exemplification of the present invention. The  
CC polypeptides, the polynucleotides, antibodies and compositions of  
CC the invention can be used as vaccines, as diagnostic reagents, and as  
CC immunogenic compositions. The polypeptides can be used in the  
CC manufacture of medicaments for treating or preventing infection due to  
CC *Neisseria* bacteria (e.g. meningitis and septicemia), to detect the  
CC presence of *Neisseria* bacteria, or to raise antibodies. They may also  
CC be used to screen for agonists or antagonists, which may themselves  
CC have use as antibacterial agents. The polynucleotides of the invention  
CC may also be used in gene therapy protocols.

XX Sequence 133 AA;

Query Match 8.5%; Score 222; DB 21; Length 133;

Best Local Similarity 42.7%; Pred. No. 7.4e-12;

Matches 44; Conservative 14; Mismatches 45; Indels 0; Gaps 0;

Db 398 GKATYDSCAACHQPDGKGVNAPPLANSPLYLNADHARAASIVANGSGKITVNGNOYE 457

Qy 30 GQKYYESNCVACHGKKGEGRTMFPPLYSDFIMKKQVLLSHVKGINGTIVNGKTYN 89

Db 458 SVMPAIALSDQOIANVITTYLNSFGNKGQLSADVDVAKAKKTK 500

Qy 90 GFWPATIAISDADIAAVALTYIMNAPDNGGSGVTEKDVQAKKKK 132

RESULT 10  
AAM48440  
ID AAM48440 standard; Protein; 1209 AA.

XX AAM48440;

XX 16-JUL-2002 (first entry)

DE 2-keto-D-gluconate dehydrogenase.

XX Cell membrane bound; 2-keto-D-gluconate dehydrogenase; enzyme;

KW 2,5-diketo-D-gluconate.

XX Unidentified.

OS Key Location/Qualifiers

XX FT Misc-difference 740 /note="Encoded by TG"

XX KR2000019366-A.

XX 06-APR-2000.

XX 10-SEP-1998; 98KR-0037413.

XX 10-SEP-1998; 98KR-0037413.

XX (KOAD) KOREA ADV INST SCI & TECHNOLOGY.

XX Shin YC, Bahn JG, Yeom DY;

XX WPI; 2001-088033/10.

XX N-PSDB; ABL52917.

XX Base sequence of novel cell membrane-bound 2-keto-D-gluconate

XX dehydrogenase gene and method for conversion into

XX 2,5-diketo-D-gluconate - Noabstract

XX Claim 3; Page 9-11; 16pp; Korean.

XX The present invention relates to a novel cell membrane-bound

XX 2-keto-D-gluconate dehydrogenase and a method for conversion into

XX 2,5-diketo-D-gluconate. The present sequence is the protein sequence for

CC the 2-keto-D-gluconate dehydrogenase.

XX Sequence 1209 AA;

Query Match  
Best Local Similarity 23.2%; Score 172; DB 22; Length 1209;  
Matches 97; Conservative 53; Mismatches 153; Indels 116; Gaps 20;

154 EASFTAPGHTSTFS--FKALQPLVYHCAVAP---VGMH-----IANGMYG---- 195  
DB QACHTAPGATATFSGGVATASPVGVIYTNITPSAGCIPYSBAERSQAVRHGVADGAG 839  
QY 136 LILVEKGLPYVDKE-----YYWQDGFYTKKYEGGLO-PEPMKAI-----RE 241  
DB 840 LTPAPAPYTSYSKITDEDLHALYYFMHGVKPVQKRRQTSLEPPFYLRFSEFMNLMFAD 899  
QY 242 DAEVY-----VFNGSV-----GALTGENALKAVGETVALLFVANGCP 278  
DB 900 DPEYISDDSGASBWRNGNYLVNGLAHCTCTPRGVLMQEGNRPLAGAPIGSWY--AP 956  
QY 279 NLTSFHVIGEIFDKVFEFGKGENHIQTTLIPAGAAITFEKVDPGVYLVDAIFR 338  
DB 957 NITS-----DAISGIGWRNDELVO-----YIKTGRA--E 984  
QY 339 AFNKALGILK-VEGEEHETIYSHKQTDVAVLPEGAP--QAIDTQEPAPKTPAPANLQEOI 395  
DB 985 GGNQAAGMAEVAEHSIQYLSDSDLKAIYVLYKSTTPIRDEGDTCPAYSGFKPADVENSI 1044  
QY 396 KA-----GKATYDSNCAACHOPDGKGVPR-APPPLANSYILNADARAASITVA 442  
DB 1045 RGRNANNANHSLTNGALFSGNCASCHQPDGASANQAPSLFPHNT-ATGMHNPANLIAA 1103  
QY 443 -----NGLSGKITFVNGNQYESVMPAIALSDQIANTVITYTLNSFGNGKGLSADDA 494  
DB 1104 ILFGVQRNTAAQGVLMPPG--FSSPSYVDKLSDAQVADISNFWLAHAGNEEYVSAGDVA 1160

# RESULT 11

AA39355  
ID AAR39355 standard; Protein; 478 AA.

XX AAR39355;

DT 13-JUN-1994 (first entry)

XX Cytochrome C.

KW C-553; improved production; prodn.; oxidative fermentation.

OS Gluconobacter suboxydans IFO 12528.

XX Key Location/Qualifiers

FT Peptide 1..36 /note= "signal peptide"

FT Peptide 37..316 /note= "mature peptide"

PN JP05049480-A.

PD 02-MAR-1993.

PF 14-JUN-1991; 91JP-0238579.

PR 14-JUN-1990; 90JP-0154096.

XX (ASAH ) ASAH CHEM IND CO LTD.

DR WPI; 1993-169634/21.

DR N-PSDB; AAQ48234.

PT Cytochrome C gene - comprises specified sequence of 1473 bases  
PT which is inserted in host cells and is used to produce cytochrome C  
PT by oxidative fermentation

XX Claim 1; Page 13-15; 20pp; Japanese.

CC The sequence is that of encoded by the cytochrome C gene from  
CC Gluconobacter suboxydans IFO 12528 which may be used to give  
CC improved prodn. of cytochrome C by oxidative fermentation.

XX Sequence 478 AA;

Query Match  
Best Local Similarity 36.0%; Score 163; DB 14; Length 478;  
Matches 49; Conservative 18; Mismatches 59; Indels 10; Gaps 6;

QY 369 LPEGAPQAIIDTQEPAPANLQEOIKAGKATYDSNCAACHQPDGKGVNAPPLA-NS 427  
DB 306 VEGEKNLG--QDDGQTALANKGGQGNAGAVYLHNCALICMNDGTGNRMPPLAGNP 362  
QY 428 DYLANDHAPASIVANGSGKITVNGNOYESV-MPAIA--LSDQIANTVITYTLNSFGNK 484  
DB 363 VITDDPSTLANVAVFG--GILPTNSAPSAVAMPGFKNHLSDOEMADVNFMRKGMGN 420  
QY 485 G-GOLSAADVAKAKKT 499  
DB 421 APGTVASDIOQLRTT 436

# RESULT 12

AA35988  
ID AAB35988 standard; Protein; 478 AA.

XX AAB35988;

DT 01-MAR-2001 (first entry)

DE Sorbitol dehydrogenase subunit 2 amino acid sequence.

XX Sorbitol dehydrogenase; SDH; Gluconobacter oxydans; D-sorbitol;

KW L-sorbose production; 2-keto-L-gulononic acid.

XX Gluconobacter oxydans.

PN WO200065066-A1.

PD 02-NOV-2000.

PF 23-APR-1999; 99WO-1B00736.

PR 23-APR-1999; 99WO-1B00736.

XX (CHOI/) CHOI E.

PA (RHEE/) RHEE S.

XX (LEEB/) LEE E.

PI Choi E, Rhee S, Lee E;

DR WPI: 2000-687351/67.

DR N-PSDB; AAC83154.

PS Claim 6; Fig 8; 96pp; English.

CC This invention relates to an isolated membrane-bound sorbitol  
CC dehydrogenase (SDH) from Gluconobacter oxydans. The invention includes  
CC polynucleotides (AAC83153 - AAC83154) encoding the 3 subunits of SDH  
CC (AAB35987 - AAB35989). Also included in the invention are two  
CC polynucleotide sequences AAC83156 and AAC83157 which contain the subunit  
CC coding regions. The sorbitol dehydrogenase polynucleotide sequences are  
CC useful for producing L-sorbose from D-sorbitol and for increasing the  
CC production of 2-keto-L-gulononic acid by transforming a host cell,  
CC especially Gluconobacter with the DNA and selecting the transformed host

CC cell. Note: An additional 66 DNA sequences (fragments of the SDH subunit  
CC genes) encoding fragments of SDH are specifically claimed, however these  
CC sequences are not given separately in the specification but are included  
CC in sequences AAC83156 and AAC83157.

SO Sequence 478 AA;

Query Match 6.2%; Score 163; DB 21; Length 478;  
Best Local Similarity 37.3%; Pred. No. 1e-05;  
Matches 50; Conservative 17; Mismatches 57; Indels 10; Gaps 6;

QY 369 LPEGAPQALDTQAPKTPAPAPNLEQIKAGKATYDSNCAACHPDGKGVNAPPLANS 428  
DB 306 VPEBKNG--QDFGKATALLKAGCKDAGAEVYLHNCALCHMDGTGVNMFPEPLAGNP 362  
QY 429 YLNDHARA-ASTVANGSLGKITNGQYBSV-MPALA--LSDQIANVITYTLNLSFGNK 484  
DB 363 VITDNTATSMANIVTFG--GILPPTNTPASVAMPFRDHLSDQIADVNVNFRKSWGNQ 420  
QY 485 G-GQLSADVDVAKAK 497  
DB 421 APGTLSSASDIRKLR 434

RESULT 13  
AAR14306  
ID AAR14306 standard; Protein; 579 AA.

AC AAR14306;

DT 10-JAN-1992 (first entry)

DE Ascorbate oxidase.

KM Enzyme.

OS Cucurbita moschata.

PN JP03210182-A.

PD 13-SEP-1991.

PF 11-JAN-1990; 90JP-0005270.

PR 11-JAN-1990; 90JP-0005270.

PA (TOYM) TOYOBO KK.

DR WPI; 1991-314587/43.

DR N-PSDB; AAQ14178.

PT DNA encoding polypeptide with ascorbate oxidase activity - derived  
PT from Cucurbita moschata, used to produce recombinant enzyme in high  
PT yield.

PS Disclosure; Fig 1; 8pp; Japanese.

CC The sequence was deduced from a DNA sequence obtd. from a clone  
CC isolated from a library prepd. from RNA extracted from the fruit  
CC of C. moschata. The DNA can be used to produce the enzyme in high  
CC yields.

SO Sequence 579 AA;

Query Match 6.2%; Score 163; DB 12; Length 579;  
Best Local Similarity 24.6%; Pred. No. 1.4e-05;  
Matches 83; Conservative 44; Mismatches 130; Indels 80; Gaps 19;

QY 101 VEVGFMT-----FCGQVPGQMTRVREGDTIEVQFSN--HPDSKMPHNVDFHATG 148

DB 40 VETVFMAPDCNENIWMGINGQFPPTIRANAGDTVVELINKLHTEGVVH--WHQILQ 96

QY 149 PG-----GGAASFTT--PGHTSTFSFKALQGLVYVHCAPVGMHIANGMGLIVER 201

DB 97 RGRPMADGTASISQCAINCEFTFFVNFVDNPGTFEYH---GILGQBSAGLYGSLIVDP 153  
QY 202 KEELPK---VDKEYVYMOQDFYTKKXGEQ-GIQPFDMKAR--EDAEVYVNG----- 250  
DB 154 PQGKKEPFHYDGEINILISDWCQSHKQEVGS---SKPLRWIGEPQITLLNGGQDP 209  
QY 251 -SYGALTGEN--ALKAKVGETVRLFVNGCGPNLTSSPHYIGELIFDKVHEGKGXENHIQ 307  
DB 210 CSTAAKYDSNLEPCCKKGESEPCAPYI-----FHMVPKTYIRI-----AS 250  
QY 308 TTLIPAGAAITE--FKVDVPGDYV---LVDAHIFRANKGALGILVEGSENEHY- 359  
DB 251 TTLAALNFAIGNHPLLVVEADGNVQPFYTSIDIDYSGESYVLLTTPQNPSENVWSV 310  
QY 360 ----SHKQPD----AVYLPEGAPQALDTQEAFTPA 387  
DB 311 GTRGRHPNTPPGILTLLNLPNSYSK-LPTSPPETPA 346

RESULT 14  
AAR32288  
ID AAR32288 standard; Protein; 478 AA.

AC AAR32288;

DT 25-MAR-2003 (updated)

DT 08-JUN-1993 (first entry)

DE Sequence of cytochrome c-553 (CO).

KM Cytochrome; oxidative fermentation; electron transfer system.

OS Gluconobacter suboxydans IFO 12528.

PN Key Location/Qualifiers

FT Peptide 1..36

FT Region 37..478

PN FR2677665-A1.

PD 18-DEC-1992.

PF 12-JUN-1992; 92FR-0007131.

PR 14-JUN-1991; 91JP-0238579.

PA (ASAH) ASAH KASEI KOGYO KK.

DR Shimizu T, Takeda Y;

DR WPI; 1993-054531/07.

DR N-PSDB; AAQ36687.

PT Plasmid contg. gene for cytochrome C-553 (CO) - and transformed  
PT cells useful in high productivity oxidative fermentation, e.g. of  
PT ethanol to acetic acid

PS Disclosure; Fig 1; 39pp; French.

CC Purified cytochrome c-553 (CO) was isolation from Gluconobacter

CC suboxydans IFO 1258 and partially sequenced. From this information,

CC 32P-labelled oligonucleotide probes were constructed and used to

CC screen a bank of genomic. EcoRI fragments from IFO 1258 in vector

CC 15pb fragment which was sequenced. This sequence lacks the N-terminal

CC region which was located from the same source and used to construct

CC a complete gene. The probes used are based on the N-terminal amino

CC acid sequence (KGMGNNA) of peptide I which was isolated following

CC limited hydrolysis of c-553(CO) using arginyl endopeptidase.

CC (updated on 25-MAR-2003 to correct PN field.)

SO Sequence 478 AA;



QY 234 DMEKAIRED-----AEYVENGSGVALLTGENA-----LK 262  
 DB 99 ALKGIARDGAHLVPAMPYTAYSIEDTDIHALVYFMHGVAPLRQDNPKTELKFPENIR 158  
 QY 263 AKYGETVRLVVG-----NGENLTSSF---HYIGEIPKRVHPEGGKNNENIOTTLL----- 311  
 DB 159 AMMISWMLLPAGPPAKGDPTQYSTIERGHYLDAL-----GHGCTCHTPRNPLMGERS 212  
 QY 312 -----PAGGAITEFKVDVGDYVLVDHALFPAFNKALG----- 346  
 DB 213 SSAYLGCTPLAGWYAPNITPSNMSGIDWSEDDLVQYLRGSGVGRQAAGMNGEAVEHS 272  
 QY 347 ILKVEGENHEIYSHKQTDVAVLPEGAPOAIDTQ-----EAPKTPAP 388  
 DB 273 FSKLTEDDLHAL-----AAVIRQ-IPKIEDSQAQPRDRFGVAVQPIVDLQKPKLDRE 324  
 QY 389 ANIOEQIKAGKATYDSNCAACHOPDGKVPNAPPEPLANSVYL-----NADHAPAASTIVANG 444  
 DB 325 DDLFPM--DGERIYVNNCAACHGIDGAGAADHFTPSLSNAVVGAPADNLIMA--IVNG 380  
 QY 445 LSGKITVNGNYESVMPAIA-----LSDQIANTVITYTLNSFGNKGGLSADDDVAKAK 497  
 DB 381 VDR--TTNG--HHVLMGPGFPTSDVQRLSDTVAKLTVVSGTFGSGDHHVTAQDVAVAR 436  
 QY 498 KTRP 501  
 DB 437 EGGP 440

RESULT 17  
 AAR12339  
 ID AAR12339 standard; Protein; 555 AA.  
 XX AAR12339;

XX 30-AUG-1991 (first entry)  
 DE Acorbate oxidase from *Cucumis sativus*.  
 XX  
 KM Acorbate oxidase; transformant; polypeptide; activity.  
 XX  
 OS *Cucumis sativus*.  
 XX  
 PN JP03108485-A.  
 XX  
 PD 08-MAY-1991.  
 XX  
 PF 19-SEP-1989; 89JP-0244257.  
 XX  
 PR 19-SEP-1989; 89JP-0244257.  
 XX  
 PA (TOYM) TOYOBO KK.  
 PA (NAGS) NAGASE SANGYO KPL.  
 XX  
 DR WPI; 1991-180924/25.  
 DR N-PSDB; AAQ12068.  
 XX  
 PT DNA encoding polypeptide having ascorbate oxidase activity -  
 PT derived from *Cucumis sativus*, and transformants producing  
 PT polypeptide in high yields  
 XX  
 PS Disclosure; Fig 2; 9pp; Japanese.  
 CC  
 CC The DNA is derived from *Cucumis sativus* and can be used to obtain  
 CC large amts. of polypeptide having ascorbate activity.  
 CC  
 XX  
 SQ Sequence 555 AA;

Query Match 5.6%; Score 146.5; DB 12; Length 555;  
 Best Local Similarity 32.5%; Pred. No. 0.00039;  
 Matches 40; Conservative 16; Mismatches 42; Indels 25; Gaps 6;  
 QY 101 VEYQFWT-----FGGQVPGQMIRVBEGDITIEVQFSN--HPDSKMPHNVDFAATG 148

DB 13 VEYMFWSPCVENIVMGINGEFFPGPTIRANAGDIVVVELTNKLTGCVIYH---MHGILQ 69  
 QY 149 PG-----GGAEASFTR--RGHTSTSPKALQPLLYRHCAVAVPGHINANGYGLIVER 201  
 DB 70 RGTPMADGTASISQCAINPEFTTYRFVVDKAGTYEYH---GHLGMQRSAGLGLSLVDP 126  
 QY 202 KEG 204  
 DB 127 PEG 129

RESULT 18  
 AAR20193  
 ID AAR20193 standard; Protein; 467 AA.

XX AAR20193;  
 AC  
 XX  
 DT 16-APR-1992 (first entry)  
 XX  
 DE ADH complex protein (mol.wt. 44.000).  
 XX  
 KM Alcohol dehydrogenase; acetic acid; fermentation.  
 XX  
 OS Acetobacter alioacetigenes NH-24.  
 XX  
 PN JP03266988-A.  
 XX  
 PD 27-NOV-1991.  
 XX  
 PF 26-MAR-1990; 90JP-0073440.  
 XX  
 PR 26-FEB-1990; 90JP-0042301.  
 PR 26-MAR-1990; 90JP-0073440.  
 XX  
 XX (NAKA-) NAKANO SUTEN KK.  
 PA  
 PA  
 DR WPI; 1992-019325/03.  
 DR N-PSDB; AAQ20384.  
 XX  
 XX  
 PT Alcohol dehydrogenase complex structural gene - used in plasmid  
 PT and enhancing efficiency of acetic acid fermentation for  
 PT transformed acetic acid bacteria  
 XX  
 PS Disclosure; Fig 4(1-3); 21pp; Japanese.  
 CC  
 CC Acetobacter transformed with the sequence encoding this protein can  
 CC enhance the efficiency of acetic acid fermentation. The ADH complex  
 CC can be easily extracted from the bacteria and purified and it can be  
 CC used for the determination of an alcohol.  
 CC See also AAQ20383-84, and -86-88.  
 CC  
 XX  
 SQ Sequence 467 AA;

Query Match 5.6%; Score 146; DB 13; Length 467;  
 Best Local Similarity 25.2%; Pred. No. 0.00033;  
 Matches 84; Conservative 50; Mismatches 127; Indels 72; Gaps 19;

QY 184 PVGMHIANGYGLIL-VEPEKGLPKVDKEYVYVNGDPYTKGYGEGQL-----OPFM-EK 237  
 DB 149 FLSMRWPLGIWRMFPSPKEDFTPAFGTDBEIRAGDYLVTGP-GHCGACHTPPGFAMQEK 207  
 QY 238 AIREDAEYVFNQSVGALTGEMALKAKVGETVR-----LPVGNQGNPLTSSFNHYIGELFDK 293  
 DB 208 AL--DAA-----GGPDLFGGAPIDNMVAPSLNDDPVVGLGRWSEDDIYTFLLSGRI DHS 260  
 QY 294 VHPGKGNNHNIQTLIPAGGAITEFKVDVGDYVLVD--HAIFPAFNKALGILKYE 351  
 DB 261 AVF-GMGD-----VYAWSTQYFTDDDLHALAK-VYKSLPVPVPSQ 299  
 QY 352 GSEENHEIYSHKQTDVAVLPEGAPOAIDTQAPKTPAPANLQEQIKAGKATYDSNCAACHQ 411  
 DB 300 GN-----YTYDPESTANMLASG-----NTASVP-----GADTYVVECAICHR 335

QY 412 PGGKVPNAPFPPLANSDYLNADHARA-ASIVANGSGKITVNGQYBSV-MPAI--ALSD 467  
 DB 336 NNGGGVARMFPPLAGNPVVVTENPTSLVNVIAHG--GVLPPSNMWSAVAMPQYSKSLSA 393  
 QY 468 QOIANTVITLNSFGNKG--GQLSADVDYAKAKKT 499  
 DB 394 QOIADVNFRTISWGNKAPGTVAADVTIKLDR 426

## RESULT 19

AAW08380  
 ID AAW08380 standard; Protein; 559 AA.  
 XX  
 AC AAW08380;  
 XX  
 DT 07-APR-1997 (first entry)  
 XX

DE Brassica napus microspore-specific clone L10 product.

XX Male sterile plant; hybrid seed; pollen; microspore; oilseed;  
 KW canola; rape; antisense RNA; stress resistance;  
 XX herbicide resistance; transgenic plant; crop protection.

OS Brassica napus.

PN WQ9640950-A1.

PD 19-DEC-1996.

XX 07-JUN-1996; 96WO-US08692.

XX 07-JUN-1995; 95US-0476864.

PA (PION-) PIONEER HI-BRED INT INC.

XX Albani DJ, Arnison PG, Fabijanski SF, Laurian RS;

DR WPI; 1997-052340/05.

DR N-PSDB; AAT48866.

PT Recombinant molecule for preparation of male sterile plants - used  
 XX to produce stress resistant plants and for increase in seed yield  
 PS Example 8; Fig 3b(1-4); 224pp; English.

CC A polypeptide sequence (AAW08380) was deduced from the coding  
 CC region contained in a fragment (AAT48863) of Brassica napus  
 CC microspore specific clone L10. Promoter regions of the L10 gene  
 CC have been used to develop pollen-specific vectors useful for  
 CC controlling antisense gene sequences for the prodn. of male  
 CC sterile plants and hybrid seed.

SQ Sequence 559 AA;

Query Match 5.6%; Score 146; DB 18; Length 559;  
 Best Local Similarity 26.1%; Pred. No. 0.00043;  
 Matches 71; Conservative 37; Mismatches 132; Indels 32; Gaps 12;

QY 100 GVEYGFMTGGGVPQGMIRREGDTEVQFSNHPD-----SKMHNVDFFHAATGPGGG 152  
 DB 43 GVPQVILINGQFPFGHNINSTNNNVIYVFNLDLPFLITWNGIQRKNCWODGTPTGM 102  
 QY 153 AEASTAPGHTSTFSFKAL-QPGLYYHCAVAVNGHIANGMVGLIVPEKBLP---K 207  
 DB 103 CP---IMPGNTYHFPQPDQIGSYFY---PTTGNRAAGGAGGGRVNSRLITIPVYAD 156  
 QY 208 VDEKYYVMGDFYTKGKYGEGLOPF-DNEKAI-REDAEYVFNQSVGALTGENA--LKA 263  
 DB 157 PEDDTYVLJGDWYTK---SHTQKRLDGRITGRPDG--IYVNGSKGSDADLFTL 211  
 QY 264 KYGTEVRLFVNGGPNLTSSFHVIGEIFDKVHEGGKGNHNIQTLLIPAG--GAITE 320

DB 212 KEGKTVRIRICNVGVKTSINFRIOHNRKMLVENEGSHVLQNDYDSLDEVEGCGFTIVTA 271  
 QY 321 FKVDVPGDYLVLDHAFRAFNKAGLITKVEG 352  
 DB 272 NQ--EPKDYVMVASSRFLKTVITTTGLRYBG 301

## RESULT 20

AAR13994  
 ID AAR13994 standard; Protein; 468 AA.  
 XX  
 AC AAR13994;  
 XX  
 DT 25-MAR-2003 (updated)  
 DT 09-DEC-1991 (first entry)  
 XX

DE A.altoacetigenes membrane-bound ADH 44kD sub-unit.

XX alcohol dehydrogenase complex; carboxylic acid production;  
 KW cytochrome c; ss.

OS Acetobacter altoacetigenes.

PN EP448969-A.

PD 02-OCT-1991.

XX 26-FEB-1991; 91EP-0102793.

XX 26-MAR-1990; 90JP-0073440.

XX 26-FEB-1990; 90JP-0042391.

PA (NAKA-) NAKANO VINEGAR CO LTD.

XX (NAKA-) NAKANO VINEGAR CO LTD.

PI Tamaki T, Takemura H, Tayama K, Fukaya M, Okumura H;

DR WPI; 1991-289462/40.

DR N-PSDB; AAQ13581.

PT Gene for membrane-bound alcohol dehydrogenase complex - obt.  
 XX from Acetobacter altoacetigenes, used for prodn. of enzyme for  
 PT converting alcohol to acid  
 PS Disclosure; Fig 4; 36pp; English.

CC This sequence was deduced from the nucleotide sequence isolated  
 CC from A.altoacetigenes MH-24 total DNA. The 44kD subunit of the  
 CC membrane-bound ADH complex is a cytochrome c.  
 CC See also AAQ13580 and AAQ13582-4.  
 CC (Updated on 25-MAR-2003 to correct PA field.)

SQ Sequence 468 AA;

Query Match 5.5%; Score 145.5; DB 12; Length 468;  
 Best Local Similarity 36.4%; Pred. No. 0.00037;  
 Matches 39; Conservative 21; Mismatches 40; Indels 7; Gaps 5;

QY 398 GKATYDSNCAQCHQDQGVNAPFPPLANSDYLNADHARA-ASIVANGSGKITVNGNQY 456  
 DB 322 GADTYVKECALIHRNDGGVARMFPPLAGNPVVVTENPTSLVNVIAHG--GVLPPSNMAP 379

QY 457 BSV-MPAI--ALSDQIANTVITLNSFGNKG--GQLSADVDYAKAKKT 499  
 DB 380 SAVAMPQYSKLSAQQIADVNFRTISWGNKAPGTVAADVTIKLDR 426

## RESULT 21

AAR06518  
 ID AAR06518 standard; Protein; 559 AA.  
 XX

AC AAR06518;





RESULT 23  
AAR72328  
ID AAR72328 standard; Protein; 529 AA

CC sequenced to reveal a novel gene, RSIac3 (given in AQ086526),  
CC encoding a laccase (MAR72328) optimally active at pH 6.0-8.5.  
CC Recombinant laccase was expressed in *E. coli*.  
CC (Updated on 25-MAR-2003 to correct PN field.)  
xx  
SQ Sequence 529 AA;

```

QY 59 LPVLDALVTNAREVPPVDRDHPAKVYVK--MEIVEKMKRLADGVEYQPMTFGGQVPCQM 116
Db 8 LPLAAVST-----PAPFAARVNYKFDIKNNVNAAPDGFORSIVSNGVLVPPTL 54
QY 117 IRVREGPTIEVOFSNH---PDSKMPHNVDH----AATPGCGGAASFTAP-----GHTS 164
Db 55 ITANKGDTLRINVTNQLTDPDSMKRAITIIHMGLEFOATTADBEDGPAFVYQCPIAQLNLTSTY 114
QY 165 TFSFKALQPLGVYHCAVAPVGMHIANGMYG-LILVEB---KSGLPKYDKREYVVMQDF 219
Db 115 EIPLRG-QGTGMVYH---AHLASQVVDGLRGPLVIYDPNDHPKSRVYDVIDDASTVYMLBDM 170
QY 220 Y-----TKGKYEGG---LQPDFMEALIFEDDAAYVFNCSVCAITGENMLKAKVGETV 269
Db 171 YHTPAVLEKMFSTNTTALLSPVDSGLINSGKRYV-----GGPAVRSAVINVRGRKY 225
QY 270 RLFPVGGNGPNLTSSFHYIGELFDKVFHEGGGKGEHNHIOQTLLIPAGG-----AAITRFKVD 324
Db 226 RLPVFNKSVGSGSTFST-----PFSH--TITPFDGIIHOLIAVNSRQIV 268

```

QY 325 VPDYVLVDHALFRAPNKGALGILKVEGEENHEIYSHKQTAANYLPRGAPALITQF--- 381  
 Dp 269 AGORYSVIVEANQTAANYWIRAPMTVAGAGTANLDPNTVFAVLHYEGAPNAEPTEGGS 328  
 QY 382 -----APKTPAANIQEOIKAKATYDS-----NCAACHOPDQKV 417  
 Dp 329 AIGTALVEENLHALINPGAPGSAADADYSLNIALIGRSTVDILRFTFNINIKYEA----- 383  
 QY 418 PNAFPLANSDYLMNDHRAASIVANGSGKITVNGNGYESWPMALMSDQIANVITYT 477  
 Dp 384 --SLPTL-----LKITLNNASNDADFTPNETHIVLPNNKVIELNITGADHP 428  
 QY 478 LNSFGKGGQGLSADPVAKAKKTKPN 502  
 Dp 429 IHLHG-----VFDIVSLGSTPN 447

RESULT 24	
AAW16301	
ID	AAW16301 standard; Protein; 529 AA.
XX	
XX	
AC	AAW16301;
XX	
DT	07-AUG-1997 (first entry)
XX	
DE	Rhizoctonia solani laccase isozyme 4.
XX	
KM	Blue copper oxidase; laccase; enzyme engineering;
KM	protein engineering; lignin depolymerisation; dye oxidation
XX	
OS	Rhizoctonia solani.
XX	
PH	Location/Qualifiers
FT	427
FT	/note= "Type I copper site ligand"
FT	470
FT	/note= "possible Type I copper site ligand"
FT	480
FT	/note= "Type I copper site ligand"
FT	485
FT	/note= "Type I copper site ligand"
FT	217..226
FT	Peptide







XX 18-DEC-2000; 2000EP-0127688.  
XX 16-DEC-1999; 99JP-0377484.  
XX 07-APR-2000; 2000JP-0159162.  
XX 03-AUG-2000; 2000JP-0280988.  
XX (KYOW ) KYOWA HAKKO KOGYO KK.  
XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;  
XX Tateishi N, Senoh A, Ikeda M, Ozaki A;  
XX WPI; 2001-316931/40.  
XX N-PSDB; AAH68245.  
XX Novel polynucleotides derived from Corynebacterium bacteria, for identifying  
XX mutation point of a gene, measuring expression of a gene, analysing  
XX expression profile or pattern of a gene and identifying homologous gene  
XX  
XX Claim 17; SEQ ID NO: 6780; 246bp + Sequence Listing; English.  
XX  
XX The present invention provides a number of nucleotide and protein  
XX sequences from the Corynebacterium Corynebacterium glutamicum. These  
XX are useful for identifying the mutation point of a gene derived from a  
XX mutant of corynebacterium bacterium, measuring expression amount and  
XX analysing the expression profile or expression pattern of a gene derived  
XX from corynebacterium bacterium, and identifying a homologue of a gene derived  
XX from corynebacterium bacterium. Corynebacterium bacteria are useful for producing  
XX amino acids, nucleic acids, vitamins, saccharides and organic acids,  
XX particularly L-lysine. The present sequence is a protein described  
XX in the exemplification of the invention.  
XX Note: The sequence data for this patent did not form part of the printed  
XX specification, but was obtained in electronic format directly from the  
XX European Patent Office.  
XX  
XX Sequence 493 AA;  
XX  
XX Query Match 5.4%; Score 140.5; DB 22; Length 493;  
XX Best Local Similarity 19.0%; Pred. No. 0.0011;  
XX Matches 80; Conservative 57; Mismatches 174; Indels 109; Gaps 15;  
XX  
XX 15 LSALMLSGCSNQADKAAQPKSSTVDAAKTANADNANASOEHGELPVIDAIVTHAEVPP 74  
XX 12 LGLVLAG-----TGAVAACTSDPGPAAS-----APG 38  
XX  
XX 75 PYDRDHPAKVYVMEVVEKML-----ADGVEYQFTRGGQVPGQMIRVREGDTIEVQ 128  
XX 39 PSLRPTPTALGEPTVRRLTLRPLSLDIGIEAKTWGVSPTGDAAIEATGADVLOVD 98  
XX  
XX 129 FSNH-PDSKMP--HNYDFHAATGPGGAASFTAPGHTSFSPKALQPLGVYVHCAVAPV 185  
XX 99 ITNELPESTSIHMHGIALHNAADGVGMQODPIEPESBSYVEVPHGGTYFPH---SHT 155  
XX  
XX 186 GMIHANGMTGLILVEKEGLPKVDKEYYVMQGDFTYKKGXGEOLOPFDMEX----- 237  
XX 156 GQLDRGLHAPLIRIPQDAEDQDVEMTIVLDWVD---GIGTDDDELKLTGWSGD 211  
XX  
XX 238 -----AIRDAEYVYVNGSVGAL-----TGENALKAKVGETVRL-FV 273  
XX  
XX 212 HNGRMGMGHHGMMHGTTPRVLGGDVGDVWYPHYLLNGRIIPRARTFEARPGDKARLRF 271  
XX  
XX 274 GNGGPNLTSFHYI--GEIPDKVHFGGKGENNIOTTLIPAGGAITERKVVDP---GD 328  
XX 272 NSGGDTI---FKVALGCHRTVTHTDGFVQPMETESIYSWGE-----RVDEVIILGD 322  
XX  
XX 329 YVLVDHAIIFRAFNKAGLILKVGEEENHE-----IYSHKOTDAVYLRBEGAP 374  
XX 323 GTFPLTALAIVGKODRAFAVIRTAGGAPARPDVDFPELSTGLLSSLKADRALRLBEGRP 382  
XX  
XX RESULT 29  
XX AABI6365

ID AABI6365 standard; Protein; 326 AA.  
XX AC AABI6365;  
XX  
XX 24-OCT-2000 (first entry)  
XX  
XX Eucalyptus grandis diphenol oxidase protein sequence SEQ ID NO:291.  
XX  
XX Plant; lignin biosynthetic pathway; Eucalyptus grandis;  
XX Pinus radiata; Monterey pine.  
XX  
XX Eucalyptus grandis.  
XX  
XX WO200022099-A1.  
XX  
XX 20-APR-2000.  
XX  
XX 06-OCT-1999; 99MO-NZ00168.  
XX  
XX 09-OCT-1998; 98US-0169789.  
XX 14-JUL-1999; 99US-0143811.  
XX  
XX (GENE-) GENESIS RES & DEV CORP LTD.  
XX (FLET-) FLETCHER CHALLENGE FORESTS LTD.  
XX  
XX Blokeberg LN, Havukkala IU;  
XX  
XX WPI; 2000-317962/27.  
XX  
XX Novel polynucleotide encoding enzymes involved in lignin-biosynthetic  
XX pathway useful for producing transgenic plants especially eucalyptus  
XX and pine species having altered lignin content, composition and  
XX structure  
XX  
XX Claim 18; Page 151-152; 213pp; English.  
XX  
XX The present invention describes isolated polynucleotides and proteins  
XX encoding and representing the enzymes cinnamate 4-hydroxylase (C4H),  
XX coumarate 3-hydroxylase (C3H), phenolase (PNL), O-methyl transferase  
XX (OMT), cinnamyl alcohol dehydrogenase (CAD), cinnamoyl-CoA reductase  
XX (CCR), phenylalanine ammonia-lyase (PAL), 4-coumarate:CoA ligase (4CL),  
XX confertif glucosyl transferase (CGr), confertif beta-glucosidase (CBG),  
XX laccase, peroxidase, ferulate-5-hydroxylase (F5H), alpha-amylase,  
XX caffeic acid methyl transferase, caffeoyl CoA methyl transferase,  
XX coumarate CoA ligase, cytochrome P450 1XX1A, diphenol oxidase, flavanol  
XX glucosyl transferase, flavonoid hydroxylase, and isoflavone reductase,  
XX which are involved in the lignin biosynthetic pathway. The  
XX polynucleotides can be used for modulating lignin content, lignin  
XX composition and the structure of a plant, especially eucalyptus and pine  
XX species, and for modifying the activity of an enzyme involved in lignin  
XX biosynthetic pathway, and for producing a plant having altered lignin  
XX content, composition and structure. They can be used for designing probes  
XX and primers useful for detecting similar DNA and RNA sequences in any  
XX organism and for PCR amplification. The lignin content can be efficiently  
XX modified using the polynucleotides. AA67908 to AA68201 and AABI6341 to  
XX AA616449 represent polynucleotide and protein sequences used in the  
XX exemplification of the present invention.  
XX  
XX Sequence 326 AA;  
XX  
XX Query Match 5.2%; Score 136.5; DB 21; Length 326;  
XX Best Local Similarity 20.9%; Pred. No. 0.0014;  
XX Matches 71; Conservative 46; Mismatches 129; Indels 93; Gaps 14;  
XX  
XX 78 RHPAKVYVMEVVEKMLADGVEYQFMTFGGQVPGQIRVREGDTIEVQSNHDSKM 137  
XX 9 RNYTFVVMKNT-----RLCS--SKPIYVANGMPEGPTLYAREDDTVLVRVSN-----RV 57  
XX  
XX 138 PHNDVPH-----AATGGGGAASFTAPGHTSTSFAL--QPLGVYVHCAVAPV 186  
XX 58 KYNVTIHHGIRQLRTGWDGPAVYITQCP--QPGGSYVNFPTTQGTGLNHNHNI----- 112  
XX 187 MHIANGMVGLILVEKEGL---PKVDKEYYVMQGDFTYKKGXG-----EGLOPFDMEX 236

```

Db      113 LMLRALTHGAIVILPKRGVPEYFPKPKHEVVVVLGEWMSDTEGVISQAIKSLGAP----- 168
QY      237 KAIREDEYVVFNSVQALT-----GENALKAKYGEIVRLPVGNGGNLTSSPHVIGEIF 291
Db      169 ----NVSDAHTTGHPPSSNCPSQGGFTLPVESGKKYMRITINAAINBELFFKTIAGH-- 222
QY      292 DKVHFEKGKGENHNIQTLLIPAGGAITEPKVDVPGDYVLVDHAIFFRAFNKALGILKVE 351
Db      223 -----QTIIVVDATYVVKPKFTDT-----IYIAP 246
QY      352 GEENHEIYSHKQTDAYVLPFG-----APQALIDQEAPEKT 385
Db      247 GQTTNALISTDQSSGKYMVAASPFMDSPAIADNMNTATAT 285

RESULT 30
AAW43448 ID AAW43448 standard; Protein; 557 AA.
XX
XX AAW43448;
AC
XX
XX 01-JUN-1998 (first entry)
XX
XX Tobacco laccase clone pTL3 protein.
DE
XX Tobacco; laccase; oxidoreductase; probe; hybridization; sycamore;
KM transgenic plant; digestibility; forage crop; paper manufacture;
KM lignin; parasite resistance.
XX
XX Nicotiana tabacum.
OS
XX
XX Key Location/Qualifiers
FH Peptide 1..22
FT /note= "signal peptide"
FT Protein 23..557
FT /note= "mature protein"

MO9745549-A1.
XX
XX 04-DEC-1997.
PD
XX
XX 30-MAY-1997; 97WO-FR00948.
PF
XX
XX 31-MAY-1996; 96FR-0006760.
PR
XX
XX (CNRS ) CENT NAT RECH SCI.
PA
XX
XX Faye L, Gomord VM, Kiefer-Meyer MC, O'Connell A;
PI WPI; 1998-032655/03.
XX
XX N-PSDB; AAV01596.
DR
XX
XX Modifying lignin biosynthesis in plants with gene encoding laccase
PT mRNA - or its anti-sense complement, especially for crops used as
PT fodder and for paper production
XX
XX
XX Claim 19; Page 32-35; 72pp; French.
XX
XX This sequence represents a full length tobacco laccase (an oxidoreductase
CC acting on diphenols) encoded by the clone pTL-3. The complete protein
CC has a calculated molecular weight of 61.9 kD and a pI of 10.08. The
CC protein has a putative signal sequence of 22 amino acids, leaving a 535
CC amino acid protein of molecular weight 509.4 kD. The mature protein
CC contains 12 putative N-glycosylation sites (Asn-Xaa-Ser/Thr) and also
CC contains 2 potential copper ion binding sites. The gene sequence was
CC isolated from a cDNA library from tobacco leaves using a probe derived
CC from the sequence of the sycamore (Acer pseudoplatanus) laccase gene.
CC The gene can be used to transform plant cells for producing transgenic
CC plants having a reduced or altered lignin content, for improved
CC digestibility of forage crops, for paper manufacture. Plants with
CC increased lignin content may have better resistance to parasites.
XX

```

```

SQ Sequence 557 AA;
Query Match 5.2%; Score 136.5; DB 19; Length 557;
Best Local Similarity 20.5%; Pred. No. 0.0031; Mismatches 163; Indels 135; Gaps 17;
Matches 89; Conservative 48;

QY 107 TFGQVPGQMIRVREGDTIEVOFSNHPDSKMPHNVDF-----AATGGGAEAS 156
Db 47 TVNGKFPPTIYAREDDTVLVKVVNH-----VKNLSIHMHGIRQLTGMADGPAYITQCP 102
QY 157 FTAPEGHTSTFSPFAL-QPGLVYTHCAVAPVGMHIANGMTGLIVERKEGL----PKVDXE 211
Db 103 I-QPGQNVYVNFPTITQQRGTLFWHNAH-----IWLRAVGAIVILPVLVPPFPKPNHE 157
QY 212 YVMQGDFTYTKGKGEGGLQPFDMKAIRE-----PAEYVFNQSVG-----AL 255
Db 158 AVVILAEWMS-----DTEAVINDAIKSGIAPVSDAHT--NGHPEVSNCAS 204
QY 256 TGENALKAKYGEIVRLPVGNGGNLTSSPHVIGEIFDKVHFEKGKGENHNIQTLLIPAGG 315
Db 205 QGGYKLVDPGKTYMRLVAINAALNBELEFFKTIAGH-----KMTVVEVDA 247
QY 316 AATFERKVDVPGDYVLVDHAIFFRAFNKALGILKVEENHEIYSHKQTDAYVLPFG--- 372
Db 248 TYIKPKFTDT-----IYIAPGQTTNVIVTANQSSGKYMVAASPF 286
QY 373 --APQALIDQEAPEKTPAPANLQEQIKAGKATYDSNCAHQPDGKGVPAFPPLANSDYL 430
Db 287 MDAPLADVNTATATLHYSGTGG-----NSHISLSTPKNATPANTFLDSRLSL 337
QY 431 NA-----DHARAASI-----VANGLSGKITVNGNQYESVPAIALSDQOI 470
Db 338 NSKKYPAKVPKIDHSLFFTVGGINPCPTCKQNGSRVAVSNVNTFPMPTALLQAHF 397
QY 471 ANVITTYTLNLSFGKNG 485
Db 398 -----FGTKG 402

RESULT 31
AAR72318 ID AAR72318 standard; Protein; 576 AA.
XX
XX AAR72318;
AC
XX
XX 25-MAR-2003 (updated)
DT 14-OCT-1995 (first entry)
XX
XX Laccase RSLacl.
DE
XX
XX RSLacl; laccase; lignin; lignosulfonate; polymerization;
KM depolymerization.
XX
XX Rhizoctonia solani.
OS
XX
XX MO9507988-A1.
PN
XX
XX 23-MAR-1995.
PD
XX
XX 13-SEP-1994; 94WO-US10264.
PF
XX
XX 17-SEP-1993; 93US-0122230.
PR 17-SEP-1993; 93US-0122827.
PR 03-DEC-1993; 93US-0162827.
PR 22-DEC-1993; 93US-0172331.
XX
XX (NOVO ) NOVO NORDISK BIOTECH INC.
PA (NOVO ) NOVO-NORDISK AS.
XX
XX Christensen BE, Schneider P, Wahlenhner JA;
PI WPI; 1995-131356/17.
XX
XX N-PSDB; AAC086524.
DR

```



Query Match	5.1%;	Score 135;	DB 19;	Length 575;
Best Local Similarity	20.0%;	Pred. No. 0.0044;		
Matches 105;	Conservative 63;	Mismatches 176;	Indels 180;	Gaps 25;

```

0Y 86 VKMEIYETKYMRLA-----DGVEYQEMTFCGQYPGQMIIRVEDGTIEVQPSN---HPDSK 136
Db 17 VLARIYEVKLKISDSEIADPGVGRNATLVNNGCPGLPIFANKDDTLKVQKQNLTPNEMY 76
0Y 137 MEHNVDFHAA-----TGPGGCAEASF---APGHTSFTSFKA-LQPLGYVYHCAY 182
Db 77 RTTSHMHGILLQHRNDDGPF-----SFTQCPRIPIRESVETYTITPLDDGTGYWYH--- 127
0Y 183 APVGMHNIANGMYGLIIVERPE---GLPKYDKXYWY-OGDFYTKXKYGEGQLQPFMEKA 238
Db 128 SHLSGQYDGLKGPVIYIFKDPHRLRYDVDEKTVLIGMY-----HESKA 175
0Y 239 I-----REDAEYVFNQSVGALTGEN-----ALKAKGETYRLFVNGGPNLTS 282
Db 176 ILASGNITQRPAVSATINGK-GRFPDDNPEANPDLYTLTKVKKGKRYRLVINSSEIASF 234
0Y 283 SFPHVIGEIRDKHAFEGCKGENNHIQITLLPAGAAITFEKVYD----- 325
Db 235 RFSV-----EGHKV--TVIADGVSTKPYVDAPDLIAGRIDCVVEANQ 277
0Y 326 -PGDYVL-----VDNAIFRAFN--KGAL 345
Db 278 EBDTYINAPLTVPNPKTAQALLVEEDRRPYHPKPGYRKWQSVSEALI KYMNMHKHNGR 337
0Y 346 GILNKVEG-----EENHEIYS-----KHQTPAVYLPEAGAPQALDQEPKTPPAPANLQ 352
Db 338 GILSGHGILKARNIBESGHHLSHSVVKRONETTVVWDESKLPLEYPPGAACSGKRPADLV 397
0Y 393 EOIKGAKATYDSNCAACHOPDQKGVNAPPEPL-----ANSDYLANADHARAAS 439
Db 398 LDL-----TFGLNFAFGTGMWI-NGIPYESPKIPTLLKILTDEDOGVTESDPTKEH---TV 448
0Y 440 IYANGLSGKITVNGNGYESVMPRIALSDOOIANYITTLNSFEN 483
Db 449 ILPKKCKIEFNKIGNSGIPRTPEVHLHG-----TWDVQOFQFN 486

```

XX	RESULT 33
XX	AAW76313
ID	AAW76313 standard; protein; 575 AA.
XX	
AC	AAW76313;
XX	
DT	08-JAN-1999 (first entry)
XX	
DE	Rhizoctonia solani (TV) laccase protein.
XX	
XX	laccase; variant; oxidation; dye transfer inhibition; bleaching;
KW	denim; lignin modification; paper strengthening; phenol polymerisation;
XX	hair dye; waste water treatment.
XX	
OS	Rhizoctonia solani.
XX	
PN	W09838287-A1.
XX	
PD	03-SEP-1998.
XX	
PF	23-FEB-1998; 98WO-DK00070.
XX	
PR	28-FEB-1997; 97DK-0000222.
XX	
PA	(NOVO ) NOVO-NORDISK AS.
XX	
PI	Svendsen A, Xu F;
XX	
DR	WPI; 1998-495393/42.
XX	
PT	New variants of Coprinus and related laccases with increased

PT oxidation potential - or altered pH optimum, or mediator or  
PT oxygen-hydroxide ion pathways, useful for oxidation, for inhibiting  
PT dye transfer and in bleaching textiles, especially as detergent  
PT additive

PS Disclosure; Pages 129-131; 147pp; English

CC The present sequence represents a laccase protein. The specification  
CC describes active laccase variants (see AMW76582, AMW76295-99 and  
CC AMW76316-17) having increased oxidation potential, altered pH optimum  
CC altered mediator and/or altered oxygen/hydroxide ion pathway. The  
CC laccase variants are used specifically to oxidise substrates, to  
CC inhibit dye transfer, and for bleaching textiles, specifically denim.  
CC They can also be used for lignin modification, strengthening paper,  
CC polymerisation of phenols, dyeing of hair and textiles and waste  
CC water treatment.

**SQ** Sequence 575 AA;

Query Match	5.1%	Score 134	DB 19	Length 575
Best Local Similarity	20.0%	Pred. No. 0.0054		
Matches 105	Conservative 63	Mismatches 176	Indels 180	Gaps 25

QY	86	YKMEVEYUWMLA-----DGEYUFWTFEGGCVNROMIYVREGDITVEGFSN---HPSDK 136
Db	17	VLARVEVGLKISDGEIAPRDGVKRNATLVNGYRFRBLIANKGDTLKVYQNKLTNPEW 76
QY	137	MHNVDFFHAA-----TGPGGAEASF-----APGHTSTPFSFKA-IQPLUYHCAV 182
Db	77	RTTSHMHGILLQHRNADDDG---SFWTOSCPVFPRESUYUTPLDLDTGTUWH--- 127
QY	183	APVGNHINGWVGLLVBERKE--GLPKYDKUYUW--OGDFYTKGVEGEGLOPFMEKA 238
Db	128	SHLSQYVDGLRGPLVITPKDRHRLYVUDEKTYVLIGDW-----HESSQA 175
QY	239	I-----REDAEYVFNFGSVGALTGEN-----ALKAVGETVRLFVGNGBNLTS 282
Db	176	ILASGNITRORPVASATINGK--GRFPDNTPRNPDTLYTLKVRGKRYLRVLVNSGIAF 234
QY	283	SPHVIGELFDKHFEGKGENNIOTTLIPAGGALITEKQVU----- 325
Db	235	RFSV-----EGHKV--TVIADGVSTKPYQVDAFDLQAQRIDCVLEANO 277
QY	326	-PGDYVL-----VDHIFRAFN--KGA 345
Db	278	EPDWTWINAPRLTNVENKTAQALLVVEEDRYRPHRKPYRKSVSSEALLIKYNNHKHKG 337
QY	346	GILKVEG-----EENHEIYS--HKOTDAVULREPGARQALDITGEARPTPRAPANO 352
Db	338	GILSGHGLKARMIEGSHHLSHSRVVKKONETTTVMDESKVLPRLEYRGAACSKRADLV 357
QY	393	EOIKAGKATYDUSNCAACHQPDGKGVNAPRPL-----ANSDYLNADHARAAS 439
Db	398	LDL-----TFGLNFATGHMMI--NGIPRESPKIPLTLLKLTIDBDGVTESDPFTEKH--TV 448
QY	440	IVANLSGKITVNGQYESWAPALILSDQOIANVITYTLNSGN 483
Db	449	ILPKKKTLEFNKSGSIPRTIHPVLLHGH-----TWDVVOG 486

RESULT 34  
ABB91862 ID ABB91862 standard; Protein; 541 AA.  
XX  
XX AC ABB91862;  
XX  
DT 31-MAY-2002 (first entry)  
XX  
XX DE Herbicidally active polypeptide SEQ ID NO 1073  
XX  
XX KW Herbicidal; plant; agriculture; herbicide.  
XX  
OS *Arabidopsis thaliana*.



XX XX WO200210210-A2.  
XX PN  
XX 07-FEB-2002.  
XX PD  
XX 28-AUG-2001; 2001WO-EP09892.  
XX PF  
XX 28-AUG-2001; 2001WO-EP09892.  
XX PR  
XX (FARB ) BAYER AG.  
XX PA  
XX Tietjen K, Weidler M;  
XX PI  
XX WPI; 2002-269010/31.  
XX DR  
XX  
XX Identifying plant target proteins for herbicidally active compounds,  
PT comparing aligning and comparing nucleic acid or amino acid sequences  
PT from plant with nucleic acid or amino acid sequences from non-plant  
PT organisms -  
XX  
XX Claim 5; SEQ ID NO 1073; 261bp + Sequence Listing; English.  
XX PS  
XX  
XX The invention relates to identifying target proteins  
CC (ABB90790-ABB94016) for herbicidally active compounds, comprising  
CC aligning and comparing nucleic acid or amino acid sequences from plant  
CC with nucleic acid or amino acid sequences from non-plant organisms using  
CC suitable search parameters, where plant sequences having an E-value  
CC greater by a factor of 3 than the E-value of most similar non-plant  
CC sequences are selected. The polypeptides or nucleic acids encoding them  
CC are useful for identifying modulators. The identified modulators are  
CC useful as herbicides.  
XX CC  
XX SQ Sequence 541 AA;

Query Match 5.0%; Score 132.5; DB 23; Length 541;  
Best Local Similarity 22.4%; Pred. No. 0.0066;  
Matches 88; Conservative 49; Mismatches 173; Indels 83; Gaps 18;  
QY 100 GVEYQFWTGGGQVGMIRREGDTIEVQSNHPS-----KMHNVDPHAAT 147  
DB 45 GVPQVILNGQFPFPIEGCVTNNIVVINKLDEPLITNGIKQKMSWODVLTGN 104  
QY 148 GGGGAASFTAGHTSTFSFKAL-OPGLVYHCAVAPVGMHANGYGLIVEPEGL- 205  
DB 105 CP-----IQKSSWTHFOLKQDIGTAY--FASISMHASAFGLANNQGSVIF 153  
QY 206 ---PKVDEYVWQGFYTKGKGEQGLQ-PFDMEKAIREDAYVFNQSVG-ALTGENA 260  
DB 154 VPPKPDADFTLLVSDMY--KMGHKLQRLDSSRALPPDGLLNGASKGLVFTGQH- 209  
QY 261 LKAKGETRLPFVNGGPNLTSSEHVGELFDVHHEGGKGEHNIOTLL---IPAGGA 316  
DB 210 ---GKIRFRISNIGISTISNFRIQGHMTLVEEG---SHTLQEVYESLIDHVGOS 260  
QY 317 ATTEFKVDVP-GDYVVLVDHAIFFAEKKGALGLIKVEEENHEIYSHKQTDVAVLP 375  
DB 261 VTVLVTLKAPVDYFIVASTRTFKPLITTTGILSYGSKRP--SHP-----LP 312  
QY 376 AIDTQEAFTPAAPANIQEOIKAGKATYDNCACHOPDGKGVNAPPLANSDYLNADHA 435  
DB 313 HHMS-----MKQARTIRLMLTANVA--SPNPGS-----SFHYGTIP 348  
QY 436 RAASIYANG---LSKITVNGNGYESVMPALAL 465  
DB 349 NRTFVLNBRAMINGKLRITYVNVSVYNDPATPL 381

RESULT 35  
AAG90799  
ID AAG90799 standard; Protein; 511 AA.  
XX AAG90799;  
XX AC  
XX

DT 26-SEP-2001 (first entry)  
XX XX  
XX C glutamicum protein fragment SEQ ID NO: 4553.  
XX DE  
XX Coryneform bacterium; amino acid synthesis; vitamin; saccharide;  
XX KM organic acid synthesis.  
XX OS  
XX Corynebacterium glutamicum.  
XX PN  
XX EP1108790-A2.  
XX PD  
XX 20-JUN-2001.  
XX PF  
XX 18-DEC-2000; 2000EP-0127688.  
XX PR  
XX 16-DEC-1999; 99JP-0377484.  
XX PR 07-APR-2000; 2000JP-0159162.  
XX PR 03-AUG-2000; 2000JP-0280988.  
XX PA  
XX (KYOW ) KYOWA HAKKO KOGYO KK.  
XX PI  
XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;  
PI Tateishi N, Senon A, Ikeda M, Ozaki A;  
XX DR  
XX WPI; 2001-376931/40.  
XX DR N-PSDB; AAH6018.  
XX PT  
XX Novel polynucleotides derived from Coryneform bacteria, for identifying  
PT mutation point of a gene, measuring expression of a gene, analysing  
PT expression profile or pattern of a gene and identifying homologous gene  
XX PT  
XX  
XX Claim 17; SEQ ID NO: 4553; 246bp + Sequence Listing; English.  
XX PS  
XX  
XX The present invention provides a number of nucleotide and protein  
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These  
CC are useful for identifying the mutation point of a gene derived from a  
CC mutant of coryneform bacterium, measuring expression amount and  
CC analysing the expression profile or expression pattern of a gene derived  
CC from Coryneform bacterium, and identifying a homologue of a gene derived  
CC from coryneform bacterium. Coryneform bacteria are useful for producing  
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,  
CC particularly L-lysine. The present sequence is a protein described  
CC in the exemplification of the invention.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from the  
CC European Patent Office.  
XX CC  
XX SQ Sequence 511 AA;

Query Match 5.0%; Score 131.5; DB 22; Length 511;  
Best Local Similarity 22.5%; Pred. No. 0.0076;  
Matches 104; Conservative 51; Mismatches 199; Indels 109; Gaps 21;  
QY 29 KAAQPKSSTVDAATAANANAASOEHOGE---LPVIDAIVTHAEVPPVP--RDHPAK 83  
DB 11 KAGAVLAATVGAQVIVACSSSDVAGXGEPRTLP-----PPADLTREGSS 58  
QY 84 VVVKMETVEKVRLLADGVEYQFWTGGQVGMIRREGDTIEVQSNHPS----- 135  
DB 59 VHFALAEAGQGESQILPDVTTKTWGFNGTHLGLTLVKKGDVHVIVNNLDEWTVHMHG 118  
QY 136 -GMPHNVP--HAAGPGGGAASFTAPRHTSTFSFKALQPLGLVYHCAVAPVGMHANG 192  
DB 119 MRLPAIADGCPHSPFGQGTWSPWTVANDAATLTVHPTHGL-----TGLHAYRG 169  
QY 193 MYGLILVEPEKGLPKVD--KEYVV-----MQGDFYTKGKYGEQGLQPFDMEKAIRBDA 243  
DB 170 LAGMITIVE-DEATDLDLDPREYGVDDIPVLVMDHFLFLEGSGIDEDLP--DGLGLDPTT 226  
QY 244 EYVFNQSVGALTGNNALKAKVGETVRLFVNGGNGNL-----TSSFHVIG-----EIFD 292  
DB 227 ANGITNAHFDATTRVRFRVLNGSNMRFY-----NLAFSDRTFTFOVLAISGLLDEPD 280



XX termination sequence.  
XX Arabidopsis thaliana.  
XX  
XX EPI033405-A2.  
XX  
XX 06-SEP-2000.  
XX  
XX  
XX 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132407.  
PR 05-MAY-1999; 99US-0132484.  
PR 06-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135253.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.

PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148655.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.

```

PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 16-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
PR 29-OCT-1999; 99US-0162143.

```

Query Match 5.0%; Score 130; DB 21; Length 554;

Best Local Similarity 22.4%; Pred. No. 0.012; Matches 104; Conservative 59; Mismatches 184; Indels 118; Gaps 23;

```

QY 82 AKVVVME-----TVEKVRMLADGVEYQFTFGGQVPGQMIRREGDTIEVOFSN 131
DB 17 ATVVVAEDRPYFHVWNVTVGTSP--GVPOQVILINQFPGPNVNSTNNVNIIVFN 74
QY 132 HPD-----SKPHNVDFHAATPGGGAASFTAPGHTSTFSFKL-QPGLVYHCAYA 183
DB 75 NLDEPFLITWNGIOHRKNCMODGTPTMCP---IMPGTNYTHFQPKDDIGSFYYPSTA 131
QY 184 PVGNHINMGWGLIVPEKGLP-----KVDKEYVVMGDFYTKKKGEOGLOPF-DMEKA 238
DB 132 ---NHRAGGFGGLRVNSRLIIPVPYADPEDDTYVILIGWYTK--SHTQKKFLDSGRT 185
QY 239 I-REDAEYVVFNGSVGALTGENA--LKAVGETVRLFVNGGPNLTSSPHVIGEIFDKVH 295
DB 186 LGRDGS--ILINGSGKGDSDADLPFLKPGKTVRVACVGLKTSINFRIOHKLIVE 243
QY 296 FEGGKGNHNIQTLIPAG--GAIT--EFKVDVPGDYVLDHAFRANKGALGILK 349
DB 244 MEGSHVLONDYDSLIVHVGQCYGTLITANQBAK-----DYMVASRFLKSVITTTGLR 298
QY 350 VEGEENHEIYSHKQTDVAVYIPEG-----AQALDIOEAPKTPAPAN 390
DB 299 YEG-----GKPASSQLPPGPGVGMASLNOFRSFRMLTASAAPNPGSYHYGKIN 350

```

```

QY 391 LOEOLKA-----GKATYDSCAACQHPDGKGVNAPF--PLANSDYLMADHAPASI 440
DB 351 ITRITKLAVNTQGVKDKLRLALN-----GVGSHTPBETFLKLAERF----- 390
QY 441 VANGLSGKITVNGNOYESV---MPAIALSDQIINAVITYTLNSF 481
DB 391 ---GVADKVF---KYDSITDNPTEQIKIVPVLNITHRTF 428

```

# RESULT 38

ABB92295 standard; Protein; 554 AA.

ABB92295;

31-MAY-2002 (first entry)

Herbicideally active polypeptide SEQ ID NO 1506.

Herbicideal; plant; agriculture; herbicide.

Arabidopsis thaliana.

WO200210210-A2.

07-FEB-2002.

28-AUG-2001; 2001WO-EP09892.

28-AUG-2001; 2001WO-EP09892.

(PARB ) BAYER AG.

Tietjen K, Weidler M;

WPI; 2002-269010/31.

Identifying plant target proteins for herbicideally active compounds,

comprising aligning and comparing nucleic acid or amino acid sequences

from plant with nucleic acid or amino acid sequences from non-plant

organisms -

Claim 5; SEQ ID NO 1506; 261pp + Sequence Listing; English.

The invention relates to identifying target proteins

aligning and comparing nucleic acid or amino acid sequences from plant

with nucleic acid or amino acid sequences from non-plant organisms using

suitable search parameters, where plant sequences having an E-value

greater by a factor of 3 than the E-value of most similar non-plant

sequences are selected. The polypeptides or nucleic acids encoding them

are useful for identifying modulators. The identified modulators are

useful as herbicides.

Sequence 554 AA;

Query Match 5.0%; Score 130; DB 23; Length 554;

Best Local Similarity 22.4%; Pred. No. 0.012; Matches 104; Conservative 59; Mismatches 184; Indels 118; Gaps 23;

```

QY 82 AKVVVME-----TVEKVRMLADGVEYQFTFGGQVPGQMIRREGDTIEVOFSN 131
DB 17 ATVVVAEDRPYFHVWNVTVGTSP--GVPOQVILINQFPGPNVNSTNNVNIIVFN 74
QY 132 HPD-----SKPHNVDFHAATPGGGAASFTAPGHTSTFSFKL-QPGLVYHCAYA 183
DB 75 NLDEPFLITWNGIOHRKNCMODGTPTMCP---IMPGTNYTHFQPKDDIGSFYYPSTA 131
QY 184 PVGNHINMGWGLIVPEKGLP-----KVDKEYVVMGDFYTKKKGEOGLOPF-DMEKA 238
DB 132 ---NHRAGGFGGLRVNSRLIIPVPYADPEDDTYVILIGWYTK--SHTQKKFLDSGRT 185
QY 239 I-REDAEYVVFNGSVGALTGENA--LKAVGETVRLFVNGGPNLTSSPHVIGEIFDKVH 295

```

Db 186 LGPDDG--LLINKSGKGDGSDAPLFTLKPKTYRVARICVNGLKTSLNFRIOHMKLVE 243  
Qy 296 FEFGKGNHNIOTLLIPAG---GAAT---EFKVDVPGDVLVDHAIIFRANFGALGILK 349  
Db 244 MEGSHVLQNDYDLDVHVGOCYGTIITANGEAR-----DYVMASRFLKSVTTTGLR 298  
Qy 350 VEGEENHEIYSHKQDAVYLPEG-----APQAIIDTQEARPTAPAN 390  
Db 299 YEG-----GKPPASSQLPPGPVGWAMSLNQFRSPRWLTFASARPNOGSYHVKIN 350  
Qy 391 LGEQITR-----GKATDSNCAACHQDQGVNAPF--PLANSDIYINADHARAASI 440  
Db 351 IRTTILVNTQGVKDKLRALN-----GVSHDPEPLKAEYF----- 390  
Qy 441 VANGLSGITVNGNOYESV---MPAIALSDQIANYITLTNSF 481  
Db 391 ---GVADKVF---KIDSTIDNPTPEQISIKIVPNVLTHTRTF 428

## RESULT 39

AAB03533 standard; Protein; 2183 AA.

AAB03533;

09-OCT-2000 (first entry)

Murine factor V SEQ ID NO:5.

XX Murine; factor V; FV; activated protein C; APC; anticoagulant;  
XX activated protein C resistant factor V; thrombosis; screening;  
XX thrombophilia.

Mus sp.

US6066778-A.

23-MAY-2000.

06-NOV-1996; 96US-0746111.

06-NOV-1996; 96US-0746111.

(UNMI ) UNIV MICHIGAN.

Ginsburg D, Cui J;

WPI; 2000-410682/35.

N-PSDB; AAA60446.

XX New transgenic mice expressing activated protein C resistant factor V  
XX as models for human thrombophilia and as models for testing in utero  
XX gene therapy protocols -

Example 1; Fig 2; 76pp; English.

XX The present invention describes transgenic mice (I) and (II) containing  
XX modifications in the factor V gene, where (I) expresses an activated  
XX protein C (APC) resistant factor V and (II) lacks the ability to express  
XX wild-type factor V. The transgenic animals (I) and (II) are useful for  
XX screening compounds with anticoagulant activity. Methods from the present  
XX invention, and the transgenic animals, are also useful in providing  
XX models for human thrombophilia. These models are useful in providing  
XX insight into the basic regulatory mechanisms of blood coagulation and  
XX pathogenesis of human thrombosis. In addition, factor V null transgenic  
XX mice, especially pregnant females may be used as a model system to test  
XX in utero gene replacement therapy protocols. The present sequence  
XX represents murine factor V, which is used in an example from the present  
XX invention.

Sequence 2183 AA;

Query Match 5.0%; Score 130; DB 21; Length 2183;  
Best Local Similarity 20.7%; Pred. No. 0.091;  
Matches 70; Conservative 46; Mismatches 120; Indels 102; Gaps 16;

Qy 63 DAIVTHAEVPPVDR-----DHPAAVYVVKMEVEKVMRLADGEVQFWT 107  
Db 357 EYIWNAPVPIPANMDIYRSOHLDNFSNOIGKRYKKVIROYEEFFTRKTNDPSIK--- 413  
Qy 108 FGQVVGQVQIRREGGTTIVQPSN---HPDSKMPHNVDFHAATGPGGAASFTAPGHTS 164  
Db 414 -QSGILGPVIRACVQRTDLKIVKRNASRPYSIYPHCVTSPYE--DGINSSTSSHT- 468  
Qy 165 TFSFKALQPG-LVYV-----HCAVAP-----VGMHIANGMYGLILVEPK 202  
Db 469 --TIRPVQGEFFTYKMNILFEDEPTENDAQCLTRYYSDDVDTRDIASGLIGLLICKS 526  
Qy 203 EGLPK-----VDKEYVYMQGDF-YTKGKRYGEGLOPF--DNEKAIREDAEVYVFNQSVG 253  
Db 527 RSLDQGVQVQVADIEQAVFAVFDEKSWYIEDNINIKFCENPDEVKDDPKF----- 578  
Qy 254 ALTGEMALKAKYGETVRLFLVNGGPRULTSSFVIGEIFD-----KVMF 296  
Db 579 -----YESNINSTINGIV-----PESISTLGCDFDPTVQWHPCSVTHDDILTIHF 624  
Qy 297 EGGR--GENHNIOTLLIPAGGAITEFPKVDVPGDYVL 331  
Db 625 TGHSFYIGRRHEDTLTLFPMRGSVTV-VMADVNTGTMWL 661

## RESULT 40

ABP78252 standard; Protein; 439 AA.

ABP78252;

07-MAR-2003 (first entry)

N. gonorrhoeae amino acid sequence SEQ ID 3034.

Antibacterial; infection; vaccine; gene therapy.

Neisseria gonorrhoeae.

MO200279243-A2.

10-OCT-2002.

12-FEB-2002; 2002WO-1B02069.

12-FEB-2001; 2001GB-0003424.

(CHIR-) CHIRON SPA.

Fontana MR, Pizsa M, Massignani V, Monaci E;

WPI; 2003-058415/05.

N-PSDB; ABZ39222.

XX New protein from Neisseria gonorrhoeae, useful for the manufacture of a  
XX medicament for treating or preventing N. gonorrhoeae infection -  
XX disclosure; Page 410; 815pp; English.

XX The present invention relates to proteins from Neisseria gonorrhoeae.  
XX Also disclosed are the nucleic acid molecules encoding the proteins and  
XX antibodies that specifically bind to the proteins. The composition  
XX comprising the protein, nucleic acid or antibody is useful for the  
XX manufacture of a medicament for treating or preventing N. gonorrhoeae  
XX infection, this may be in the form of a vaccine or gene therapy.  
XX Sequences given in records ABP76736-ABP81046 represent nucleic acid  
XX molecules of the invention.

SQ Sequence 439 AA;  
Query Match 4.9%; Score 129.5; DB 24; Length 439;  
Best Local Similarity 22.6%; Pred. No. 0.0092;  
Matches 79; Conservative 47; Mismatches 131; Indels 93; Gaps 18;  
QY 184 PVGNHIANGMVGLIVE-----PKEGLPKVDEKYYVMOD-----FYTGRK----- 224  
133 POKQIAQNLFTNYCICQCHGSDAKSGKGFNLTDSDWLMGDDPKIHETIEKGRVATMPA 192  
QY 225 -----YGEQGLQPFMEKAIREDAYV-FNGSVGALTGENALKAKV---GETVRLFYGNG 276  
193 WGPALGEEG-----VKDVAHYVMSLSKPKQYDEERAARQALFSGPPANCFTCHG 243  
QY 277 -----GPNLTSFPHVIGEIPDKVHFEGGKGEN-----HNIQTLLIPAGGAITEFK 322  
244 DKGGGIGLGNLT-----DQVWLMGGTOKSIIETITNGRSSQMPAMGHFLDKX 293  
QY 323 VDVPGDYVLVDHAFRAFNK-GALGILKVEGEEHETYSKQTDVYLLPEGAPQALDTQE 381  
294 LHIWTAVV-----WGLSNKDGKAPVKKAEPAPAPAPAPAPAAQAASEAKPAAAEPK 347  
QY 382 APKTPAPANLOEQIKA-GKATYDSCAACHQPDGKGVNAPPLANDYLNADHAPAST 440  
348 AEEKAAPA-----AKADGKQYETVCAACH---GNAIPGIPIHVGTKADW--ADRIK----- 393  
QY 441 VANGLSGKITVNGNOYE--SVMPAIA-----LSDQAIANVITYTLNSFGNK 484  
394 -----KGGDTLHKHAIIEGFNTMPAKGGRGLSDDEVAADVIMVNGSGK 438  
Db

Search completed: August 27, 2003, 18:29:33  
Job time : 88 secs

*Handwritten:* 9/14/1999 Page 1

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: September 5, 2003, 05:17:53 ; Search time 407 Seconds

(without alignments)  
3329.528 Million cell updates/sec

Title: US-10-088-045-2

Sequence: 1 MSKPTLIKTTTLCALSLML.....NKGQLSADVAKAKTKKEN 502

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Delop 6.0 , Delext 7.0

Searched: 2552756 segs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODE=frame+ p2n model -DRV=xlh  
-Q=/cgr2\_1/USFTO.spool/US10088045/runcat\_04092003\_083141\_3908/app\_query.fasta\_1.647  
-DB=N.Geneseq.15Jun03 -QFMT=fastap -SUFTX=ring -MINMATCH=0.1 -LOOPL=0  
-LOOPEXT=0 -UNITS=bits -STRAT=1 -END=-1 -MATRIX=biosum62 -TRANS=human40.cdi  
-LIST=5 -DOCALLIGN=200 -THR SCORE=500 -THR MAX=100 -THR MIN=0 -ALIGN=40  
-MODE=LOCAL -OUTFMT=ptc -NORML=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US10088045 @CCN 1.1.312 @runcat\_04092003\_083141\_3908 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEOUTERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGIOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGADOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N.Geneseq.15Jun03.\*  
1: /SIDSI/gcgdata4/geneseq/geneseqn-emb1/NA1980.DAT.\*  
2: /SIDSI/gcgdata4/geneseq/geneseqn-emb1/NA1981.DAT.\*  
3: /SIDSI/gcgdata4/geneseq/geneseqn-emb1/NA1982.DAT.\*  
4: /SIDSI/gcgdata4/geneseq/geneseqn-emb1/NA1983.DAT.\*  
5: /SIDSI/gcgdata4/geneseq/geneseqn-emb1/NA1984.DAT.\*  
6: /SIDSI/gcgdata4/geneseq/geneseqn-emb1/NA1985.DAT.\*  
7: /SIDSI/gcgdata4/geneseq/geneseqn-emb1/NA1986.DAT.\*  
8: /SIDSI/gcgdata4/geneseq/geneseqn-emb1/NA1987.DAT.\*  
9: /SIDSI/gcgdata4/geneseq/geneseqn-emb1/NA1988.DAT.\*  
10: /SIDSI/gcgdata4/geneseq/geneseqn-emb1/NA1989.DAT.\*  
11: /SIDSI/gcgdata4/geneseq/geneseqn-emb1/NA1990.DAT.\*  
12: /SIDSI/gcgdata4/geneseq/geneseqn-emb1/NA1991.DAT.\*  
13: /SIDSI/gcgdata4/geneseq/geneseqn-emb1/NA1992.DAT.\*  
14: /SIDSI/gcgdata4/geneseq/geneseqn-emb1/NA1993.DAT.\*  
15: /SIDSI/gcgdata4/geneseq/geneseqn-emb1/NA1994.DAT.\*  
16: /SIDSI/gcgdata4/geneseq/geneseqn-emb1/NA1995.DAT.\*  
17: /SIDSI/gcgdata4/geneseq/geneseqn-emb1/NA1996.DAT.\*  
18: /SIDSI/gcgdata4/geneseq/geneseqn-emb1/NA1997.DAT.\*  
19: /SIDSI/gcgdata4/geneseq/geneseqn-emb1/NA1998.DAT.\*  
20: /SIDSI/gcgdata4/geneseq/geneseqn-emb1/NA1999.DAT.\*  
21: /SIDSI/gcgdata4/geneseq/geneseqn-emb1/NA2000.DAT.\*  
22: /SIDSI/gcgdata4/geneseq/geneseqn-emb1/NA2001A.DAT.\*  
23: /SIDSI/gcgdata4/geneseq/geneseqn-emb1/NA2001B.DAT.\*  
24: /SIDSI/gcgdata4/geneseq/geneseqn-emb1/NA2002.DAT.\*  
25: /SIDSI/gcgdata4/geneseq/geneseqn-emb1/NA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed.

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2626	100.0	1506	22	AAFS5657	Nucleotide sequenc
2	2626	100.0	1509	22	AAFS5658	Nucleotide sequenc
3	2626	100.0	66986	22	AAFT28542	Genomic fragment #
4	1394	53.1	349980	21	AAFT21611	Neisseria meningit
5	1394	53.1	837096	21	AAAB1489	N. meningitidis pa
6	1388	52.9	1176	25	AB238597	N. gonorrhoeae nuc
7	1346.5	51.3	7919	21	AAAB1536	N. meningitidis pa
8	1034	39.4	691	24	AAI46493	M. catarrhalis MCAI
9	349	13.3	37337	23	AAFS5918	Propionibacterium
10	237	9.0	399	21	AAZ53036	Neisseria gonorrhoe
11	237	9.0	552	25	AB239100	N. gonorrhoeae nuc
12	222	8.5	396	21	AAZ53038	Neisseria meningit
13	222	8.5	402	21	AAZ53037	Neisseria meningit
14	222	8.5	12893	21	AAA81733	N. meningitidis pa
15	222	8.5	349980	21	AAFT21608	Neisseria meningit
16	222	8.5	1437668	21	AAAB1490	N. meningitidis B
17	184	7.0	341	21	AAZ53034	Neisseria meningit
18	184	7.0	341	21	AAZ53035	Neisseria meningit
19	175	6.7	4290	23	ABL52917	2-Keto-D-glucosate
20	172	6.5	342	21	AAZ53033	Neisseria gonorrhoe
21	164	6.2	1437	14	AAO36687	Sequence encoding
22	163	6.2	1437	14	AAO48234	Cytochrome C gene.
23	163	6.2	1437	20	ABN96893	Glucosyltransferase
24	163	6.2	1437	21	AAAC83154	Sorbitol dehydroge
25	163	6.2	2064	12	AAO14178	Gene encoding asco
26	163	6.2	4830	20	ABN96894	Glucosyltransferase
27	163	6.2	4830	21	AAAC83156	DNA encoding Gluco
28	162	6.2	2007	22	AAFT26419	Pseudomonas sp hea
29	158.5	6.0	4403765	22	AAI99683	Mycobacterium tube
30	158.5	6.0	4411529	22	AAI99682	G. oxydans D-sordi
31	155	5.9	5187	20	AAZ57912	Pseudomonas aerugi
32	152.5	5.8	1125	25	ABT14589	ADH complex struct
33	152	5.7	1404	13	AAO20384	ADH complex struct
34	150	5.7	1404	12	AAO13581	A.alcacetigenes me
35	146.5	5.6	1668	12	AAQ12068	Ascorbate oxidase
36	142	5.4	1533	22	AAH66018	C glutamicum codin
37	142	5.4	1614	22	AAFT71220	Corynebacterium g1
38	142	5.4	349980	22	AAH68527	C glutamicum codin
39	141	5.4	1588	18	AAZ63317	Rhizoctonia solani
40	141	5.4	1588	21	AAZ24235	R. solani lactase
41	141	5.4	1672	16	AAQ86526	Lactase gene RSLac
42	140.5	5.4	1479	22	AAH68245	C glutamicum codin
43	140.5	5.4	309400	22	AAH68534	C glutamicum codin
44	139	5.3	1722	24	ABT13870	Arbidopsis thalia
45	136.5	5.2	980	21	AAA68115	Eucalyptus grandis

ALIGNMENTS

RESULT 1	AAFS5657	standard; DNA; 1506 BP.
ID	AAFS5657	
XX	AAFS5657	
AC	AAFS5657	
XX	AAFS5657	
DT	11-JUN-2001	(first entry) :
XX	AAFS5657	
DE	Nucleotide sequence of a BASB109 polypeptide.	
XX	AAFS5657	
KM	BASB109; bacterial infection; vaccine; genetic immunisation; ss.	
XX	AAFS5657	
OS	Moraxella catarrhalis.	
XX	AAFS5657	
Key	Location/Qualifiers	
FT	1..1506	
CDS	/*tag= a	

```

FT      /product= "BASB109"
FT      /note= "no termination codon given"
FT
FT
FT

```

PN WO200119996-A1.

PD 22-MAR-2001

14 SEP 2000; 2000WO=EP09035.

PR 14-SEP-1999; 99GB-0021691.

PA (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.

Thonnard J;  
PI  
yy

WPI; 2001-257883/26  
F-132DB; AVE7669

XX  
DT  
Novel  
BASCPI09  
polym

PT Novel Basb109 polypeptides of *Moraxella catarrhalis* useful for  
PT diagnostic, prophylactic and therapeutic purposes against microbiobla  
PT diseases, preferably bacterial infections -

Claim 13; Page 65-66; 93pp; English.

The present sequence encodes a BASH109 polypeptide of Moraxella catarrhalis. BASH109 polypeptides and polynucleotides are useful for treating bacterial infections, and as research reagents and materials for the treatment of and diagnosis of diseases, particularly human diseases. They are useful for inducing an immune response in an individual, and to assess the binding of small molecule substrates and ligands in, for e.g. cells, cell-free preparations, chemical libraries, and natural product mixtures. BASH109 polynucleotides are useful for therapeutic or prophylactic purposes, in particular genetic immunisation and in diagnosis of the stage and type of infection.

**SQ** Sequence 1506 BP; 431 A; 356 C; 363 G; 356 T; 0 other;

**Alignment Scores:**

Pred. No.:	2,526-227	Length:	156
Score:	2626.00	Matches:	502
Percent Similarity:	100.00%	Conservative:	0
Beet Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	22	Gaps:	0

Percent Similarity:

Query Match:

US-10-088-045-2 (1-502) X AAF55657 (1-1506)

QY	1	MeSeSerLYpPThrThLeuIleLysThrThLeuIleCYsaLaLeuSerLaLeuMeLLeu	20
Db	1	ANGTCMAACCTACTTTGATTAABAAACAACCTTAATTGTGGCTTAAGGCAATTATGCTC	60
QY	21	SerGIYCySeSerAmGlnAlaAspLYsaLaalagInProlYSeSerThValAspLa	40
Db	61	AGGGGTGTAGCAATTCMAACGGACAAGCCGCCCAAAAGACAGACAGCGTGAAGCGCT	120
QY	41	AlaAlaLYeThrAlaAsnAlaAspAsnAlaAlaSeGInGluHISgInGlyLysLeuPro	60
Db	121	GGCGCCMAACAACGAAATTCAGATTAATGCTGCTCACAAGAACATCAAGGCGAAGCTCCT	180
QY	61	ValIleAspAlaIleValIleThrHisAlaLeuProGluValProProProValAspArgAspHis	80
Db	181	GTCATTGATGACATTGTTACCATCCATCCACAGAAAGTTCACACACCTGTTGACCGTGACAC	240
QY	81	ProAlaLYeValValIleLysMeGInLysThrValGluLYsaValMeArgLysLeuAlaAspGly	100
Db	241	CCCGCCMAAGTGTGTGTAAATAATGGAAACCGTTGAAAAAGTCATGCGTTCGACGAATGGC	300
QY	101	ValGluLYrGlnPheTrpThrPheGlyGlyGlnValProGlyGlnMeLLeaArgValArg	120
Db	301	GTGGATATACAGTTTGGACATTTGGCGGCTCAAGTTCAGGGCAGATGATTCGTGTCGT	360
QY	121	GluGlyAspThrIleGluValGlnPheSerAsnHisAspLeuAspSerLYeMeLProHisAsn	140

D	36	GAAGGCGACACACATCGAAGTGCAGCTTCTCAAAACCAACCAACGATTTCAAAAATGCCCATAT	420
Q	141	ValAspPheHisalaalatrnglyProGlyGlyGlyalagluAlaSerPheThrAlaPro	160
D	421	GTTGACCTTTCACGCTGCCACAGAGGCGTCCGGCGGGGAGAAAGCGTCAATTACCCACCG	480
Q	161	GlyHisPheSerThrPheSerPheLeuAlaLeuGlnProGlyLeuTyrValTyrHisGly	180
D	481	GGTCATACATCAACCTTTAGTTTAAAGCCTTACAGCCTGGTTGTATGCTATACACGT	540
Q	181	AlaValAlaProValGlyMetHisAlaAlaAsnGlyMetTyrGlyLeuAlaLeuValGlu	200
D	541	CGCGTTGCCCTCTGTGGCATGCACATGCTCTAATGGCATGATATGGTTGATTTGGTTGA	600
Q	201	ProLySGlyGlyLeuProLySValAspLySGlyTyrTyrValMetGlnLyAspPheTyr	220
D	601	CCAAAAGAGGGCTTGCCAAAAGTGAATAAGAAATCATATGCATGCAAGCGCATTTTAT	660
Q	221	ThrLySGlyTyrGlyGlyGlnGlyLeuGlnProPheAspMetGlyValAlaIleArg	240
D	661	ACCAAAAGCAAAATATGGCGAAACAAAGGCTCAAGCCCTTGTATATGAAAAAGCCATTGA	720
Q	241	GluAspAlaGlyTyrValAlaPheAsnGlySerValGlyAlaLeuThrGlyAlaAsnAla	260
D	721	GAAGATCTGAAATATGTGCTCTTAAATGGCTTCGGTGGGGCGGTGATCTGGTGAAAAATCT	780
Q	261	LeuLySAlaLySValGlyIleThrValArgLeuPheValGlyAsnGlyGlyProAsnLeu	280
D	781	CTAAAAAGCCAAAGTTGGCGAAACGTGCTTATTTGTGGGTAAACGGCGCCCGAATTTG	840
Q	281	ThrSerSerPheHisValIleGlyGlyIlePheAspLySValHisPheGlyGlyGlyS	300
D	841	ACATCATCATTTCCATGTCATGTGGAGATTTTGTATAGATTCATTGAGGGGTGTAG	900
Q	301	GlyGlyAsnHisAsnIleGlnThrThrLeuIleProAlaGlyGlyAlaAlaIleThrGlu	320
D	901	GGTGAAAAACCAATATCCAAACCAAGCTAATCCCAAGAGTGGCGCTGCCATCACTGA	960
Q	321	PheLySValAspValProGlyAspTyrValLeuValAspHisAlaIlePheArgAlaPhe	340
D	961	TTTAAAGGTGATGTCGGGGTGTATATGTCTTGAGTACCATGTCATCTTCGGTCCCTTT	1020
Q	341	AsnLySGlyAlaLeuGlyIleLeuLySValGlyGlyGlyLeuAsnHisGlyIleTyrSer	360
D	1021	AACAAAGGGGCAATTTGGGCATCTTAAGTGAAGTGAAGTGAAGAAATCATGAAATTTATTC	1080
Q	361	HisLySGlnThrAspAlaValTyrLeuProGlyGlyAlaProGlnAlaIleAspThrGln	380
D	1081	CACAAACAAACAGACGCTGTCTTATCTGCAGAGGGGTCCCAACAGCAATTTGATACCA	1140
Q	381	GluAlaProLySThrProAlaProAlaAsnLeuGlnGlnIleLySAlaGlyValAla	400
D	1141	GAGGACCCCAAAACCTGCACCTGCACCTTCAAGACGAGATTAAAGCAGTAAAGCA	1200
Q	401	ThrTyrAspSerAsnCySAlaAlaCyHisGlnProAspGlyLySglyValProAsnAla	420
D	1201	ACCTAATAGACTTAACGTGCTGCTTGCAACCACTGATGTTAAAGCGTGCAAACCT	1260
Q	421	PheProProLeuAlaAsnSerAspTyrLeuAsnAlaAspHisAlaArgAlaAlaSerIle	440
D	1261	TTCCACCGCTTGCACATCTGACTATCTAAGCGGCAACACGCTCGTCCGCCAGCATC	1320
Q	441	ValAlaAsnGlyLeuSerGlyLySAlaThrValAsnGlyAsnGlnTyrGlySerValMet	460
D	1321	GTGGCAAAATGGAATTGTCTGTTAAGATTACCGTCAATGGCAACCAATATGAAGGTCATG	1380
Q	461	ProAlaIleAlaLeuSerAspGlnGlnIleAlaAsnValIleThrTyrThrLeuAsnSer	480
D	1381	CTGCAGATTGCTCTGAGCGACCAACAGATTGCCAATGTCTACCTACACGCTTAACGCG	1440
Q	481	PheGlyAsnLySGlyGlyGlnLeuSerAlaAspAspValAlaLySAlaLySValThrLyS	500
D	1441	TTTGTGTAACAAAGGCGGTCAACTCAGTGCAGACGATGTGGCAAAAGCCCAAAAACCAAG	1500

AAAT	420
AAPro	160
ACCG	480
Acys	180
ALGTGT	540
Glu	200
GGNA	600
peTyr	220
TATAT	660
ARG	240
CCGA	720
ALA	260
GGCT	780
Ieu	280
TTTG	840
Llys	300
AAAG	900
Glu	320
GGNA	960
Phe	340
TTTTT	1020
Ser	360
TTCA	1080
Gln	380
CCA	1140
ALA	400
GGCA	1200
ALA	420
GGCT	1260
ILE	440
TATC	1320
Met	460
ATG	1380
Ser	480
AAAGC	1440
Llys	500
AAAG	1500



QY 501 Proben 502  
DB 1501 CCAAC 1506

RESULT 2  
ID AAF55658 standard; DNA; 1509 BP.  
AC AAF55658;  
XX  
XX  
XX 11-JUN-2001 (first entry)  
DE Nucleotide sequence of a BASB109 polypeptide.  
XX  
XX BASB109; bacterial infection; vaccine; genetic immunisation; ss.  
XX  
XX Moraxella catarrhalis.

Key Location/Qualifiers  
FH 1..1509  
FT CDS /tag= a  
FT /product= "BASB109"

MOZ00T19996-41.  
XX  
XX 22-MAR-2001.  
XX  
XX 14-SEP-2000; 2000WO-EP09035.  
XX  
XX 14-SEP-1999; 99GB-0021691.  
XX  
XX (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.  
XX  
XX Thonnard J;  
XX  
XX WPI; 2001-257883/26.  
XX  
XX P-PSDB; AAB67668.

PT Novel BASB109 polypeptides of Moraxella catarrhalis useful for  
PT diagnostic, prophylactic and therapeutic purposes against microbial  
PT diseases, preferably bacterial infections  
XX  
XX  
XX Class 13; Page 65; 93pp; English.

CC The present sequence encodes a BASB109 polypeptide of Moraxella  
CC catarrhalis. BASB109 polypeptides and polynucleotides are useful for  
CC treating bacterial infections, and as research reagents and materials  
CC for the treatment of and diagnosis of diseases, particularly human  
CC diseases. They are useful for inducing an immune response in an  
CC individual, and to assess the binding of small molecule substrates and  
CC ligands in, for e.g. cells, cell-free preparations, chemical libraries,  
CC and natural product mixtures. BASB109 polynucleotides are useful for  
CC therapeutic or prophylactic purposes, in particular genetic immunisation  
CC and in diagnosis of the stage and type of infection.

XX  
XX Sequence 1509 BP; 432 A; 356 C; 364 G; 357 T; 0 other;  
SQ

Alignment Scores:  
Pred. No.: 2.53e-227 Length: 1509  
Score: 2626.00 Matches: 502  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 22 Gaps: 0

US-10-088-045-2 (1-502) x AAF55658 (1-1509)

QY 1 MetSerLeuProthThreuIleuThrIleuIleCysAlaIeuSerAlaIeuMetLeu 20  
DB 1 AGTCTTAAGGCTTACTTGAATTAACAACCTTAATTTGCTTAAGTGATGATGCTC 60

QY 21 SerGlyCysSerAsnGlnIaAspLysAlaIaGlnProIysSerSerThrValaAspAla 40

DB 61 AGTGGTTGATCAATCAAGCGGCAAGCGCCCAAAAAGCAGCGGTAGACGCT 120  
QY 41 AlaAlaLysThrAlaAsnAlaAspAsnAlaAlaSerGlnGluHISGlnGlyLeuPro 60  
DB 121 GCCGCCAAGACGCAAAATGCAATATGCTGCTTCAAGAACATCAACGCAAGCTGCT 180  
QY 61 ValIleAspAlaIleValThrHisAlaProGluValaProProValaAspArgAspHis 80  
DB 181 GTCATTGATGCAATGTTAGCGATGCACAGAAATGCCACCTGTTGACCGTGAACAC 240  
QY 81 ProAlaLysValaValaLysMetGluThrValaGluLysValaMetArgLeuAlaAspGly 100  
DB 241 CCCGCCAAAGTGCTGTAATAAATGGAACCCGTGAAAAAGTATGCTGTCAGATGAC 300  
QY 101 ValGluIyrGlnPheThrPheGlyGlnValaProGluGlnMetIleArgValaArg 120  
DB 301 GTGAATATCAAGTTTGGACATTGCGGTCAGATTCCAGGCGAATGATTGTGTGCT 360  
QY 121 GluGlyAspThrIleGluValaGlnPheSerAsnHisProAspSerLysMetProHisAsn 140  
DB 361 GAAGCGACACCATGCAAGTGCAGTTCTCAACACCAACCATTTCAAAAATGCCCATAT 420  
QY 141 ValaAspPheHisAlaAlaThrGlyProGlyGlyAlaGluAlaSerPheThrAlaPro 160  
DB 421 GTTGACTTTCACGCTGCACAGGGCGGCGGCGGAGAAAGCGTCATTATCCGACCG 480  
QY 161 GlyHisThrSerThrPheSerPheLysAlaLeuGlnProGlyLeuThrValaIyrHisCys 180  
DB 481 GGTCAATACATCAACTTGAATTTAAGCTTACAGCTGTTGTGATGTATCATCTGT 540  
QY 181 AlaValaIaProValaGlyMetHisIleAlaAsnGlyMetGlyLeuIleLeuValaGlu 200  
DB 541 GCGGTTGCCCTGTTGGCATGCATGCTGTATGCAATGATAGTATGTTGTTGTTGA 600  
QY 201 ProLysGluGlyLeuProGlyValaLysGlyIyrValaMetGlnGlyAspPheIyr 220  
DB 601 CCAAAAAGAGGGCTGCCAAAAGTAGATAAGAACTACTATGTCATCAAGGCGACCTTAT 660  
QY 221 ThrIyrGlyLysGlyGlyGlnGlnIyrLeuGlnProPheAspMetGlnLysAlaIleArg 240  
DB 661 ACCAAAGGCAATATAGCGCAACAGGTCTACAGCCCTTTGAATGGAATAAACCATTCGA 720  
QY 241 GluAspAlaGluIyrValaValaPheAsnGlySerValaGlyAlaLeuThrGlyLysAla 260  
DB 721 GAAGATGCTGAATATGTTCTTAAATGTTGCGGGGCGCTTACTGCTGTAATAATGCT 780  
QY 261 LeuLysAlaLysValaGlyIyrValaValaArgLeuPheValaGlyAsnGlyIyrProAsnLeu 280  
DB 781 CTAAAGGCCAAGGTTGGGAAACTGTGCTTATTTGTGGTAAACGGCGCCGAATTTG 840  
QY 281 ThrSerSerPheHisValaIleGlyGluIlePheAspLysValaHisPheGlnGlyLys 300  
DB 841 ACATCATATCTCCATGATGTTGATGATGATTTTGAATGATGATGATGATGATGATG 900  
QY 301 GlyIuAsnHisAsnIleGlnThrThrLeuIleProAlaGlyAlaAlaIleThrGlu 320  
DB 901 GGTGAATAACCAATATTCAAACCAACGCTAATCCAGACAGGTGGCGCTGCCACTGA 960  
QY 321 PheLysValaAspValaProGlyAspIyrValaLeuValaAspHisAlaIlePheArgAlaPhe 340  
DB 961 TTTAAGGTGAGATGTCGGGATATTAATGCTGTTGACCAATGCCATCTTCGTCGCTTT 1020  
QY 341 AsnLysGlyAlaLeuGlyIleLeuLysValaGluGlyGluGluAsnHisGlnLysIleTyrSer 360  
DB 1021 AACCAAGGGGCAATGGGATATCTTAAGGTGAAGGTGAAGAAATCAATGAGATTAATCA 1080  
QY 361 HisLysGlnThrAspAlaValaIyrLeuProGluGlyAlaProGluAlaIleAspThrGln 380  
DB 1081 CACAAACAAACAGAGCTGTCTATCTGCCAAGAGGTGCCCAACAGCAATATATACCCAA 1140  
QY 381 GluAlaProLysThrProAlaProAlaAsnLeuGlnGlnIleLysAlaGlyLysAla 400

D3 1141 GAAGCAACCAAAACCTGCACTGCACTTCAACAGAGATTAAAGGTAAGCA 1200  
 QY 401 Th-TyrAspSerAsnGlyValAlaCysHisGlnProAspGlyValProAsnAla 420  
 Db 1201 ACCTATACCTTACTTACTGCTGCTTGTCCACCACTTATGTAAGCGTCCAAACGCT 1260  
 QY 421 PheProIleuValAsnSerAspTyrLeuAsnAlaAspHisAlaArgAlaAsnIle 440  
 Db 1261 TTCACACCGCTTGGCAACTGCACTATCTGAGACCGACACGCTGCTGCGACGATC 1320  
 QY 441 ValAlaAsnGlyLeuSerGlyValIleThrValAsnGlyAsnGlnTyrGluSerValMet 460  
 Db 1321 GTGGCAAAATGATGTCGTGTAAGATTACCGTCATGCAACCAATATGAAGCGTCATG 1380  
 QY 461 ProAlaIleAlaLeuSerAspGlnGlnIleAlaAsnValIleThrTyrThrLeuAsnSer 480  
 Db 1381 CCGCGATGCTCTGACGACCAACAGATTGCCATCTCATCTACAGCTTAACAGC 1440  
 QY 481 PheGlyAsnIleGlyGlnLeuSerAlaAspAspValAlaValAlaValSlyThrLys 500  
 Db 1441 TTGGTAAACAAGCGGCTCAACTCAGTCAGACGATGTGGCAAAAGCCAAAGCAAG 1500  
 QY 501 ProAsn 502  
 Db 1501 CCAAAAC 1506  
 RESULT 3  
 AAF28542  
 ID AAF28542 standard; DNA; 66986 BP.  
 AC AAF28542;  
 DT 04-APR-2001 (first entry)  
 XX Genomic fragment #29.  
 DE Genomic  
 KM Genomic library; bacteria; human upper airway; otitis media; sinusitis;  
 KW bronchopulmonary; endocarditis; meningitis; ss.  
 XX Moraxella catarrhalis.  
 OS  
 PN WO200028966-A2  
 PD 28-DEC-2000  
 PF 16-JUN-2000; 2000MO-US16649.  
 XX 18-JUN-1999; 99US-0140121.  
 PR (INCYTE GENOMICS INC.  
 PA Lagace RE, Patterson C, Berg KL;  
 PI WPI; 2001-041427/05.  
 XX Genomic library for identifying diagnostic and therapeutic  
 PT compositions, and for identifying virulence factors, regulatory  
 PT elements, and drug targets; comprises Moraxella catarrhalis nucleic  
 PT acids -  
 XX Claim 1; Page 247-263; 545BP; English.  
 PS The present invention relates to a Moraxella catarrhalis genomic library  
 CC comprising of a combination of 41 nucleic acid molecules (see  
 CC AAF28542-AAF28544). The library has a number of uses described in the  
 CC specification e.g. is useful for identifying diagnostic and therapeutic  
 CC compositions. M. catarrhalis (Branhamella catarrhalis) is a large  
 CC aerobic, gram-negative diplococcus, normally found among the bacterial  
 CC flora of human upper airways. M. catarrhalis is known to cause acute,  
 CC localised infections such as otitis media, sinusitis and bronchopulmonary  
 CC infection and life-threatening, systemic diseases including endocarditis  
 CC and meningitis.

SQ Sequence 66986 BP; 18889 A; 13427 C; 15112 G; 19558 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 4,1e-225 Length: 66986  
 Score: 2626.00 Matches: 502  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 22 Gaps: 0  
 US-10-086-045-2 (1-502) x AAF28542 (1-66986)  
 QY 1 MetSerLysProThrLeuIleLysThrThrLeuIleCysAlaLeuSerAlaLeuMetLeu 20  
 Db 3742 ATGTCTAACCTACTTGTGATTAACCACTTAATTTGGCTTAAAGTCATGATGCTC 3801  
 QY 21 SerGlyCysSerAsnGlnAlaAspLysAlaAlaGlnProLysSerSerThrValAspAla 40  
 Db 3802 AGTGGTTGTAGCAATCAACCGGCAAAAGCCGCCCAAAAGCAAGCAGGTAGACGCT 3861  
 QY 41 AlaAlaLysThrAlaAsnAlaAspAsnAlaAlaSerGlnGlnHisGlnGlyLeuPro 60  
 Db 3862 GCGCCAGACACGACAAATGCAATGATGCTGCTCAAGAAACATCAAGCGACGCTGCT 3921  
 QY 61 ValIleAspAlaIleValIleThrHisAlaProGluValProProValAspArgAspHis 80  
 Db 3922 GTCATTTGATGCCATTTGTTACGATGACCAAGATTCACCACTGTTGACCGTGAAC 3981  
 QY 81 ProAlaLysValAlaValLysMetGlyThrValGluLysValMetCysGlyLeuAlaAspGly 100  
 Db 3982 CCGGCCAAAGTGGTGTAAATAATGGAACCGTGTGAATAAGTATGCTGCGACAGTGGC 4041  
 QY 101 ValGluTyrGlnPheThrThrPheGlyGlyGlnValProGlyGlyMetIleArgValArg 120  
 Db 4042 GTGGAATATCAAGTTTGGATTTGGACATTTGGCGGTCAAGTTCCAGGCGACATGATTCGTGCT 4101  
 QY 121 GlyIleAspThrIleGluValGlnPheSerAsnHisProAspSerLysMetProHisAsn 140  
 Db 4102 GAAGGGAACCATGCAAGTGCAGTTCTCAAAACCAACCGATTCAAAATAATGCCCATTAAT 4161  
 QY 141 ValAspPheHisAlaAlaThrGlyProGlyGlyGlyAlaGluAlaSerPheThrAlaPro 160  
 Db 4162 GTTGACTTTCACGCTGCCACAGGCGCTGGCGGGCGGAAGCGCATTTACCGCACCG 4221  
 QY 161 GlyHisThrSerThrPheSerPheLysAlaLeuGlnProGlyLysThrValTyrHisCys 180  
 Db 4222 GGTCAATCAATCAACCTTTGTTTAAGCTTTACAGCTGTTTGTATGTATCACTCT 4281  
 QY 181 AlaValAlaProValGlyMetHisIleAlaAsnGlyMetTyrGlyLeuIleLeuValGlu 200  
 Db 4282 GCGGTTGCCCTGTTGGCATGCACTGCTAATGGCATGTATGGTTGATTTGGTTGA 4341  
 QY 201 ProLysGlyGlyLeuProLysValAspLysGlyTyrTyrValMetGlnGlyAspPheTyr 220  
 Db 4342 CCAAAAGAGGCGCTTGGCAAAAGTAGTAAAGATCTAATGATGCAAGCGCACTTTTAT 4401  
 QY 221 ThrLysGlyLysTyrGlyGlnGlnGlnLeuGlnProPheAspMetGluLysAlaIleArg 240  
 Db 4402 ACCAAAGGCAATATGGCGAACAAGGCTACAGCCCTTTGATATGGAAAGCCATTCCA 4461  
 QY 241 GluAspAlaGlyTyrValValAlaPheAsnGlySerValGlyAlaLeuThrGlyGluAsnAla 260  
 Db 4462 GAAGATGCTGAATATGTTCTTTTAATGTTGGTGGGCGGTGATCGTGAATAATGCT 4521  
 QY 261 LeuLysAlaLysValAlaGlyThrValArgLeuPheValGlyAsnGlyGlyProAsnLeu 280  
 Db 4522 CTAAAGGCAAGGTGGCCAAACGTGCTTATTTGGGTAACGCGCCGCAATTTTG 4581  
 QY 281 ThrSerSerPheHisValIleGlyGluIlePheAspLysValHisPheGlyGlyGlyLys 300  
 Db 4582 ACATCTCATCTCCATGTCATTTGGTGAATTTTGAATTAAGTTCACTTTAAGGGTGAAG 4641  
 QY 301 GlyLysAsnHisAsnIleGlnThrThrLeuIleProAlaGlyValAlaAlaIleThrGln 320

```

Db      4642 GGTGAAGAACCAATATCCAAACACGCTATCCAGCAGTGGCGTGCATCAGTGA 4701
Qy      321 PheLysValAspValProGlyAspTyrValLeuValAspHisAlaIlePheArgAlaPhe 340
Db      4702 TTTAAGTGGATGATGCGGATGATATGCTTGGTGGACCATGCCATCTTCGCGCTTT 4761
Qy      341 AsnLysGlyValAlaLeuGlyIleLeuLysValGluGlyLysAsnHisGluIleTyrSer 360
Db      4762 AACAAAGGGGCATTTGGGCATCTTAAAGTGAAGGTAAGAAATATCATGATTTATCA 4821
Qy      361 HisLysGlnThrAspAlaValTyrLeuProGluGlyAlaProGlnAlaIleAspThrGln 380
Db      4822 CACAAACAAACACACGCTGCTATCTGCGCAGAGGAGTCCCAACAGCAATGTGATCCCA 4881
Qy      381 GluAlaProLysThrProAlaProAlaAsnLeuGlnGluGlnIleLysAlaGlyLysAla 400
Db      4882 GAAGCACCCAAACACCTGACCTGCGCACTTACAGAGCATTAAGCAAGGTAAAGGCA 4941
Qy      401 ThrTyrAspSerAsnCysAlaAlaCysHisGlnProAspGlyLysGlyValProAsnAla 420
Db      4942 ACTATGACTCTAAGCTGCTGCTGTGTCACCAAGCTGATGTAAGCGCGCCAAAGCT 5001
Qy      421 PheProProLeuAlaAsnSerAspTyrLeuAsnAlaAspHisAlaArgAlaIleAspTyr 440
Db      5002 TTCCCAACCGCTTCCCAACTGACTGATCTGACAGCCGACACCGCTGCGCAGCATC 5061
Qy      441 ValAlaAsnGlyLeuSerGlyLysIleThrValAsnGlyLysGlnIleTyrGlySerValMet 460
Db      5062 GTGGCAAAATGATGTTGTGTGAATTAACGCTCAATGGCAACAAATATGAAGCGTCATG 5121
Qy      461 ProAlaIleLeuLeuSerAspGlnIleAlaAsnValIleThrTyrThrLeuAsnSer 480
Db      5122 CCGGCGATGCTCTGAGCGGACCAAGATGGCCAAAGTGTATCATCAGCTTAACAGC 5181
Qy      481 PheGlyAsnLysGlyGlyGlnLeuSerAlaAspAspValAlaLysAlaLysLysThrLys 500
Db      5182 TTTGGTAAACAAAGCGGCTCAACTCAGTGACAGATGTGGCAAAAGCAAAACCAAG 5241
Qy      501 ProAsn 502
Db      5242 CCNAAC 5247

```

RESULT 4  
AAF21611  
ID AAF21611 standard; DNA; 349980 BP.  
AC AAF21611;  
XX  
XX  
XX 13-MAR-2001 (first entry)  
DE Neisseria meningitidis B nucleotide sequence SEQ ID NO:112.  
XX  
XX Neisseria meningitidis; Neisseria gonorrhoeae; immunogenic; vaccine;  
KM diagnosis; antigen; detection; infection; gene therapy; antibacterial;  
XX de.  
OS Neisseria meningitidis.  
XX  
XX WO200066791-A1.  
XX  
XX 09-NOV-2000.  
XX  
XX 08-MAR-2000; 2000MO-US05928.  
XX  
XX 30-APR-1999; 99US-0132068.  
XX 08-OCT-1999; 99MO-US23573.  
XX 28-FEB-2000; 2000GB-0004695.  
XX  
XX (CHIR ) CHIRON CORP.  
XX (GENO-) INST GENOMIC RES.  
XX  
XX Plazza M, Hickey E, Peterson J, Tettein H, Venter JC, Maignani V;

PI Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V, Rappuoli R;  
PI Frazer CM, Grandi G;  
XX  
XX MPI; 2000-647603/62.  
XX  
XX Neisseria meningitidis B full length genome sequence and open reading  
PT frames are used to detect, treat and prevent Neisserial infections.  
XX  
XX Claim 7; Appendix A; 692pp; English.

CC The present invention describes the full length genome of  
CC Neisseria meningitidis B (NMB). The sequences in AAF21544 and AAF21607  
CC to AAF21613 represent fragments of the NMB genomic sequence, as the  
CC sequence was too long to go in a record on its own it was split into 8  
CC sequences which overlap each other at the beginning and end of each  
CC sequence by 49980 bp (i.e. the last 49980 bp of AAF21544 is repeated at  
CC the beginning of AAF21607, the last 49980 bp of AAF21607 are repeated at  
CC the beginning of AAF21608, and so on). AAF21545 to AAF21588 encode the  
CC Neisseria proteins given in AAB58550 to AAB58593, and AAF21589 to  
CC AAF21606 represent PCR primers which are used in the exemplification of  
CC the present invention. The NMB genome and fragments from it have  
CC antibacterial activity, and can be used in vaccines and gene therapy.  
CC Neisseria nucleic acids, proteins and/or antibodies which binds to the  
CC proteins can be used in compositions for treating or preventing infection  
CC due to Neisserial bacteria or as a diagnostic reagent for detecting the  
CC presence of Neisserial bacteria or of antibodies raised to Neisserial  
CC bacteria. Computers, computer memory, computer storage medium or computer  
CC databases can be used in a search to identify open reading frames (ORFs)  
CC or coding sequences within the NMB genome. The DNA sequences provide  
CC further opportunities to find antigenic or immunogenic proteins which are  
CC more effective in vaccines than the outer membrane proteins currently  
CC used.

SO Sequence 349980 BP; 87189 A; 93501 C; 84627 G; 84663 T; 0 other;

#### Alignment Scores:

Pred. No.: 5,66e-113 Length: 349980  
Score: 1394.04 Matches: 265  
Percent Similarity: 79.74% Conservative: 46  
Best Local Similarity: 67.95% Mismatches: 75  
Query Match: 53.08% Indels: 4  
DB: 21 Gaps: 2

US-10-088-045-2 (1-502) x AAF21611 (1-349980)

```

Qy      1 MetSerLysProThrLeuIleLysThrThrLeuIleCysAlaLeuSerAlaLeuMetLeu 20
Db      187900 TTACAAAGGAAATATTAATGAAGCCCAAGCCTTAAGTCAATGATTCCTTCTATTTC 187959
Qy      21 SerGlyCysSerAsnGlnAlaAspLysAlaIleGlnProLysSerSerThrValAspAla 40
Db      187960 GCATTAGCGCGCTGCGGCGGCGCAACCTGCGGCGCAAGCCCTGCCAAACCTCGCGCT 188019
Qy      41 AlaAlaLysThrAla---AsnAlaAspAsnAlaAlaSerGlnGlnLysGlnGlnLys 59
Db      188020 GCGGCGGAAAGCCGCAAGCTCCGCGGCAAAACCGCGGCAAAACCGCGGCAAGCTG 188079
Qy      60 ProValIleAspAlaIleValThrHisAlaProGluValProProProValAspArgAsp 79
Db      188080 CCGGTATGATGCGGTATCAACCCACGCTCCGGAAGTCCCTTCGCAATGACCGCGGAC 188139
Qy      80 HisProAlaLysValValValLysMetGluThrValGluLysValMetArgLeuAlaAsp 99
Db      188140 TACCCGCGCAAGTCCGCGTAAATGAAACCGTCGAAACCAATGACCATGGAAGAC 188199
Qy      100 GlyValGluTyrGlnPheThrPheGlyGlyGlnValProGlyGlnMetIleArgVal 119
Db      188200 GGTGTGAATACCGCTACTGACGATTTGACGCGCAGCTTCGCGGCGGTATGCGCGTA 188259
Qy      120 ArgGluGlyAspThrIleGluValGlnPheSerAsnHisProAspSerLysMetProHis 139
Db      188260 CCGGAAGCGGATGACGTTGAAAGTTCACCAACATCTTCTTCAACGTTCCGCGAC 188319

```

```

QY 140 AsnValaAspPheHisAlaAlaThrGlyProGlyGlyAlaGluAlaSerPheThrAla 159
DB 188320 AACGTGACCTTCCACGGGGGTACCGGCGCGGGCGCGGCGGCAACCTTACCGCT 188379
QY 160 ProGlyHisThrSerThrPheSerPheValAlaLeuGlnProGlyLeuThrValThrHis 179
DB 188380 CCGGGCGGTCTTCCACATTCAGCTTCAAAAGCCCTGCAACCGGGTCTGTACATCTACAC 188439
QY 180 CysAlaValAlaProValGlyMetHisAlaAlaSerGlyMetThrGlyLeuValLeuVal 199
DB 188440 TGGCCCTCCGACCGGTCTGATATCAGATCCCAACCGATATGACGATGATTTGGTC 188499
QY 200 GluProGlyGlyGlyLeuProGlyValAlaAspGlyLeuThrValMetGlnGlyAspPhe 219
DB 188500 GAGCTTAAGAAAGGCTGCGCAAGGTGATTAACAGTTTACATCGTCCAAAGGCACTTC 188559
QY 220 TyrThrGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 239
DB 188560 TACACCAAGGCAAAAGGCGCGCAAGGTCTGCAACCGTTCGATATGCAAAAGCGGTT 188619
QY 240 ArgGluAspAlaGlyValAlaValPheAsnGlySerValGlyAlaLeuThrGlyGluAsn 259
DB 188620 GCGCAACAGCTGAAATCGTCGTATTCACGCTACGTAAGTCTATCGCCGCGGATAC 188679
QY 260 AlaLeuValAlaValGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 279
DB 188680 GCGCTGAAGCCAAAGCAAGGCGAAAGTCTGATATGATGATGATGATGATGATGATGAT 188739
QY 280 LeuThrSerSerPheHisAlaAlaGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 299
DB 188740 TTGGTATCTTCTTCCACGCTATCGGGAATCTTCAACAAAGTTATGTTGAAGGCGC 188799
QY 300 LysGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 319
DB 188800 AAACGTATTAACCAAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 188859
QY 320 GluPheValAlaValAlaValAlaValAlaValAlaValAlaValAlaValAlaValAla 339
DB 188860 GAATTCAAAGTCGATCCCGGCGGCAAGTCTGTTGACCACTTATCTTCCGCGCA 188919
QY 340 PheAsnValGlyAlaLeuGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 359
DB 188920 TTCAACAAAGCGGCAAGTCTGATTAAGTGAAGTCAAGCAAGCAAGCAAGCAAGTATG 188979
QY 360 SerHisGlyGlnThrAspAlaValAlaValAlaValAlaValAlaValAlaValAlaVal 379
DB 188980 ACTCAAAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 189030
QY 380 GlnGluAlaProGlyThrProAlaProAla 389
DB 189031 GCTTCGCTCCCGGAGCTTCCGCGCA 189060

```

RESULT 5  
 AAA81489  
 ID. AAA81489 standard; DNA; 837096 BP.

```

AC AAA81489;
XX
XX
DT 04-DEC-2000 (first entry)
XX
XX N. meningitidis partial DNA sequence gnm_37 SEQ ID NO:37.
XX
XX Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;
KM antigen; vaccine; diagnosis; infection; antibacterial; identification;
KM Meningococcus B; MenB; ds.
XX
XX Neisseria meningitidis.
XX
XX OS
XX PN MO200022430-A2.
XX
XX PD 20-APR-2000.
XX
XX 08-OCT-1999; 99WO-US23573.

```

```

XX
XX 09-OCT-1998; 98US-0103794.
PR 30-APR-1999; 99US-0132068.
XX
XX (CHIR) CHIRON CORP.
PA
PI Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC,
PI Maignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;
PI Rappapoli R, Piza M;
XX
XX WPI; 2000-318079/27.
XX
XX Isolated nucleotide sequences of Neisseria meningitidis which can be
PT used in the diagnosis and treatment of N. meningitidis infection and
PT other Neisserial infections, for example, N.gonorrhoea -
XX
XX Claim 7, Page 629-865, 1760pp; English.
XX
XX The present invention describes methods of obtaining immunogenic
CC proteins from Neisseria genomic sequences. AAA81453 to AAA82414
CC represent specifically claimed Neisseria meningitidis genomic DNA
CC sequences. AAA81260 to AAA81303 and AAB25620 to AAB25663 represent
CC Neisseria DNA sequences and their corresponding proteins. AAA81254 to
CC AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the
CC isolation of Neisseria meningitidis DNA sequences; and AAA81322 to
CC AAA81452 represent Neisseria meningitidis MenB polynucleotide ORF
CC sequences, which are all used in the exemplification of the present
CC invention. The nucleic acid sequences, protein sequences, and antibodies
CC against them, can be used in the manufacture of a composition. The
CC composition can be used as a medicament (or in the manufacture of a
CC medicament) for treating, preventing or diagnosing infection due to
CC Neisserial bacteria. For example, some of the identified proteins could
CC be components of vaccines against Meningococcus B; against all serotypes;
CC and/or against all pathogenic Neisseriae. Identification of sequences
CC from the bacterium will also facilitate production of biological probes,
CC particularly organism-specific probes. Attempts to make efficacious
CC Meningococcus B vaccines have failed mainly due to antigen tolerance.
CC Multivalent vaccines have also been tried but none have successfully
CC overcome antigenic variability. The provision of further, complete
CC sequences may provide an opportunity to identify secreted or surface
CC exposed proteins that may be presumed targets for the immune system and
CC which are not antigenically variable or at least more conserved than
CC other more variable regions.
XX
XX
XX Sequence 837096 BP; 207534 A; 227065 C; 205215 G; 197280 T; 2 other;
SQ

```

Alignment Scores:  
 Pred. No.: 1,82e-112 Length: 837096  
 Score: 1394.00 Matches: 265  
 Percent Similarity: 79.74% Conservative: 46  
 Best Local Similarity: 67.95% Mismatches: 75  
 Query Match: 53.08% Indels: 4  
 DB: 21 Gaps: 2

```

US-10-088-045-2 (1-502) x AAA81489 (1-837096)
QY 1 MetSerLysProThrLeuThrLeuThrLeuThrLeuThrLeuThrLeuThrLeuThrLeu 20
DB 252671 TTACAAAGAAATTTATGAAACCGCAAGCTTATGCTGCAATGATTCCTTATTC 252730
QY 21 SerGlyCysSerAsnGlnAlaAspValAlaGlnProLysSerSerThrValAlaAla 40
DB 252731 GCATTGCGCGCTCGCGCGCGCAACTGCGCGCAAGCCCTGCGCAAACTCCTCGGCT 252790
QY 41 AlaAlaLysThrAla--AsnAlaAspAlaAlaSerGlnGluHisGlnGlyGluLeu 59
DB 252791 GCGCGCGAAGCGCAAGCTCCCGCACAACCGCGCGCAACACCTGCGCGCAAGT 252850
QY 60 ProValAlaAspAlaLeuAlaThrHisAlaProGlyValProProProValAlaAsp 79
DB 252851 CCGGTATGATGGGTGATCCACCGCTCCCGAAGTCTCTGCAATGACCCGCGAC 252910
QY 80 HisProAlaLysValAlaValAlaLysMetGluThrValGluLysValMetArgLeuAlaAsp 99

```

```

DB      252911  TACCCCGCMAAGTCGCGCTAAATAAGAAACCGTCGAAAAAACCCTGACCCATGGAAGAC 252970
OY      100  GYValAGluTyrGlnPheThrPhePheGlyGlyValProGlyGlnMetIleArgVal 119
DB      252971  GGTGTGGAATACCGCTACTGACATTTGACGGACGCTTCGGGCGGTATGATCCGGTA 253030
OY      120  ArgGluGlyAspThrIleGluValGlnPheSerAsnHisProAspSerLysMetProHis 139
DB      253031  CCGGAAGCGCATACGGTTGAAGTGAATTTTCCAAACAATCTTCTACCCGTCGGAC 253090
OY      140  AsnValAspPheHisAlaIleThrGlyProGlyGlyValAlaGluLasePheThrAla 159
DB      253091  AACGTGACTTCACACCGGCTACCGGCAAGGCGGCGCGGCCCAACCTTTACCGGT 253150
OY      160  ProGlyHisThrSerThrPheSerPheLysAlaLeuGluProGlyLeuValTyrHis 179
DB      253151  CCGGCGCGTACTTCACATTCCTTCAAGCCCTCAACCGGCTCTGATACCTACAC 253210
OY      180  CysAlaValAlaProValGlyMetHisIleAlaAsnGlyMetTyrGlyLeuIleVal 199
DB      253211  TCGGCGCGTGGACCGGTCGGTATGACATCGCAACGATATGACGCTGATTTGGTC 253270
OY      200  GluProGlyGluGlyLeuProLysValAspLysGlyTyrTyrValMetGlnGlyAspPhe 219
DB      253271  GAGCTTAAAGAAAGCGCTCGCGAAAGTGATAAAGATTCATCATCTCCAAAGCGACTTC 253330
OY      220  TyrThrLysGlyLysTyrGlyGluGlnGlyLeuGlnProPheAspMetGluLysAlaIle 239
DB      253331  TACACCAAAAGGCAAAAGGCGCGCAAGGCTGCAACCGTTCGATVGAACAAAGCGGT 253390
OY      240  ArgGluAspAlaGluTyrValValPheAsnGlySerValGlyValLeuThrGlyGluAsn 259
DB      253391  GCGGAACACCGCTGAATACGTCGATTCACACGTCACGTCAGTGGTCTATCCCGCGCATAC 253450
OY      260  AlaLeuLysAlaLysValGlyValValArgLeuPheValGlyAsnGlyGlyProAsn 279
DB      253451  GCGCTGAAAGCAAGCAAGCGCGCAACCTGATACGATGATGATGATGATGATGATGATG 253510
OY      280  LeuThrSerSerPheHisValIleGlyGluIlePheAspLysValHisPheGluGlyGly 299
DB      253511  TTGGTATCTCTCTCCACGTCATCGCGCAAAATCTTGACAAAGTTATGTTGAAGCGGCG 253570
OY      300  LysGlyGluAsnHisAsnIleGlnThrLeuIleProAlaGlyValAlaIleThr 319
DB      253571  AATCGATTAACGAATAACGTACAAACACCATGCTTCCTCGCGCGCTCTGCCATCTTC 253630
OY      320  GluPheLysValAspValProGlyAspTyrValLeuValAspHisAlaIlePheArgAla 339
DB      253631  GAATTCAAAGTCGACATCCCGGCGACCTACATTTGTTGACACATCTATCTTCGCGCA 253690
OY      340  PheAsnLysGlyAlaLeuGlyIleLeuLysValGluGlyGluLysAsnIleGluTyr 359
DB      253691  TTCACCAAAAGCGCACTGGGTCAATTGAAGTGAAGGTCGCAAAAAACCTGAAACATG 253750
OY      360  SerHisLysGlnThrAspAlaValTyrLeuProGluGlyValAspGlnAlaIleAspThr 379
DB      253751  ACTCAAAATTTAGTATACCGCTTACCGCGGTAAACGGTGACGCTCTCTGCT----- 253801
OY      380  GlnGluAlaProLysThrProAlaProAla 389
DB      253802  GCTTCGCGTCCGCGACGCTTCGCGCGGCA 253831
RESULT 6
AB238597
ID      AB238597 standard; DNA; 1176 BP.
XX
XX      AB238597;
AC
XX
XX      07-MAR-2003 (first entry)
DT
XX      N. gonorrhoeae nucleotide sequence SEQ ID 1783.
DE
XX

```

```

KW      Antibacterial; infection; vaccine; gene therapy; gene; ds.
XX
XX      Neisseria gonorrhoeae.
OS
XX      WO200279243-A2.
XX
XX      10-OCT-2002.
XX
XX      12-FEB-2002; 2002WO-1B02069.
XX
XX      12-FEB-2001; 2001GB-0003424.
XX
XX      (CHIR-) CHIRON SPA.
XX
XX      Fontana MR, Pizza M, Malignani V, Monaci E;
PI      WPI: 2003-058415/05.
DR      P-PSDB; ABP77627.
XX
XX      New protein from Neisseria gonorrhoeae, useful for the manufacture of a
PT      medicament for treating or preventing N. gonorrhoeae infection -
PS      Disclosure; Page 316; 815pp; English.
XX
XX      The present invention relates to proteins from Neisseria gonorrhoeae.
CC      Also disclosed are the nucleic acid molecules encoding the proteins and
CC      antibodies that specifically bind to the proteins. The composition
CC      comprising the protein, nucleic acid or antibody is useful for the
CC      manufacture of a medicament for treating or preventing N. gonorrhoeae
CC      infection, this may be in the form of a vaccine or gene therapy.
CC      Sequences given in records AB237706-AB242016 represent nucleic acid
XX      molecules of the invention.
XX
XX      Sequence 1176 BP; 285 A; 371 C; 299 G; 221 T; 0 other;
SQ
Alignment Scores:
Pred. No.: 9,47e-116 Length: 1176
Score: 1388.00 Matches: 265
Percent Similarity: 81.05% Conservative: 43
Best Local Similarity: 69.74% Mismatches: 66
Query Match: 52.86% Indels: 6
DB: 25 Gaps: 3
US-10-088-045-2 (1-502) x AB238597 (1-1176)
OY      11  LeuIleCysAlaLeuSerAlaLeuMetLeuSerGlyCysSerAsnGlnAlaAspLysAla 30
DB      25  ARGATTCCTTCTTATTCGCAATGGCCGCTCGGCG-----GGCGAACCTCGC 72
OY      31  AlAGlnProLysSerSerThrValAspAlaAlaIleValThrAla--AsnAlaAspAsn 49
DB      73  GCGCAAGCCCGCGCAAAACCCCTCCGCGCAAGACCGCAAGCGCAAGTTCGCCGCAACA 132
OY      50  AlaAlaSerGlnGluHisGlnGlyLeuProValIleAspAlaIleValThrHisAla 69
DB      133  GCTACCGCGCAAAACCGTCGACGGGCACTGCGCGCATGATGATGATGATGATGATGATG 192
OY      70  ProGluValProProProValAspArgAspHisPProAlaValValValValLysMetGlu 89
DB      193  CCGGAAGTACCTCCCGCAATCGACCGGCACTATCTCGCCAAAGTACGGGTAAATGGA 252
OY      90  ThrValGluLysValMetArgLeuAlaAspGlyValGluTyrGlnPheThrPheGly 109
DB      253  ACCGTGAAATAACCATGAATAATGACAGAGGGTGAATACCGCTACTGACATTTGAC 312
OY      110  GlyGluValProGlyGlnMetIleArgValArgGluGlyAspThrIleGluValGlnPhe 129
DB      313  GCGCAAGTTCGCGCGCGTATGATCGCGGTACCGGAAGGCGATACGCTGAATGCAATTC 372
OY      130  SerAsnHisProAspSerLysMetProHisAsnValAspPheHisAlaIleThrGlyPro 149
DB      373  TCCAAACAATCTTCTTACGCTTCGCGCAACGTCGACTTCACGCGCGCAACCGGTGAC 432

```

OY	150	GIYGIYGLIYALIGLIJLJLAEERPEHTRIALPROGLIYHISETHSerThrPheSerPheLys	169
Db	433	GGCGGCGGCGCAAGCCGCAACCTTTACCGCCCGGGCGGCACTTCACATTCAGCTTCAAA	492
OY	170	ALALEuGlnProGlyLeuTyValTyTrpValIleuValAlaValAlaProValGlyMetHisIle	189
Db	493	GGCTTCGCAACCGGGCTCTGATCATCTCACTCGACGCGCGCTGGCGCGGTATGCACTC	552
OY	190	ALAAANGlyMetTyGlyLeuIleLeuValGlyProLysGlnGlyLeuProLysValAsp	209
Db	553	GGCAACGGATGATGACGGTCTGATTTTGGTCGAGCCTAAAGAAAGGCTGGCGAAAGTGAT	612
OY	210	LysGlyTyTrpValMetGlnGlyAspPheTyTrpIleLysGlyLysTyTrpGlyLysGlnGly	229
Db	613	AAAGAGTTCTACATGTCGTCGCAAGGCACTTCTACACAAAGGCAAAAAGGCGCGCAAGC	672
OY	230	LeuGlnProPheAspMetLysLysValAlaIleAspGlyLysValGlyTyValValPheAsn	249
Db	673	CTGCACACGTTGATGATGACCAAGCGCTGGCGCAACGCTGAATACGCTGATTTCAAC	732
OY	250	GlySerValGlyAlaLeuThrGlyLysLysAlaLeuLysValAlaLysValGlyLysThrVal	269
Db	733	GGCCGCGTAGGGGCTATCGCGCGCGGATTAACGCGCTGAAGCAAGGCGGCAAAACCGTG	792
OY	270	ArgLeuPheValGlyAlaAsnGlyLysProAsnLeuThrSerSerPheHisValIleGlyLys	289
Db	793	CGTATGTCGTCGGATGACGGCGGCGGCAACCTTGGTGCTCTCTTCACGATCATCGGCGAA	852
OY	290	IlePheAspLysValHisAspGlnGlyLysGlyLysLysAlaAsnHisAsnIleGlnIleThrTrp	309
Db	853	ATCTTCGCAAAAGTTATGTTGTAAGGCGGCAAACTGATTAACGAAGAGCTACAAAGCAAC	912
OY	310	LeuIleProAlaGlyGlyAlaAlaAlaIleThrGlyPheLysValAspValProGlyAspTrp	329
Db	913	ATCGGCTGCGCGCGGCGGTCTGCGCACTCGTCGATTCGAAAGTGCATCCCGGCAACTAC	972
OY	330	ValLeuValAspHisAlaAlaIlePheAspTrpAlaPheAsnLysGlyAlaLeuGlyIleLeuLys	349
Db	973	ACTTTGGTCGACCACTCATCTTCGCGGCATTCACAAAGGCGGCTGGGGGAATTGAAA	1032
OY	350	ValGlyGlyLysGlyLysAsnHisGlnIleLeuGlySerHisLysGlnIleThrAspAlaValTyLeu	369
Db	1033	GTAAGGGGTGGGAAAGCCGTGAATCATGACTCAAAAATTGAGATGATCCGTTACGCC	1092
OY	370	ProGlnGlyAlaProGlnAlaAlaIleAspThrGlnGlnAlaProLysThrProAlaProAla	389
Db	1093	GGCAGCGGCGGCGGCTTCTGCC---CGTGTGCTTCCGACCGGCGGTGCTTCCCGCGCA	1149

RESULT 7  
AAA81536/c  
ID AAA81536 standard; DNA; 7919 BP  
xy

AC	AAA81536;	
XX		
DT	04-DEC-2000	(first entry)
DE	N. meningitidis partial DNA sequence gnm_83	SEQ ID NO:83.
XX		
KM	Neisseria meningitidis: Neisseria gonorrhoeae; genome; immunogenic;	
XX	antigen; vaccine; diagnosis; infection; antibacterial; identification	
KM	Meningococcus B; MenB; ds.	
OS	Neisseria meningitidis.	
XX		
PN	WO200022430-A2.	
XX		
PD	20-APR-2000.	
XX		
PF	08-OCT-1999;	99WO-US23573.
XX		
PR	09-OCT-1998;	98US-0103794.
XX		
PR	30-APR-1999;	99US-0132068.
XX		

PA (CHIR ) CHIRON CORP.  
XX  
XX  
XX Frazer CM, Hickey E, Peterson J, Tetzelin H, Vanter JC,  
PI Maizman V, Galocci C, Mora M, Ratti G, Scarselli M, Scarlato V,  
PI Rappuoli R, Piazza M,  
XX  
XX WPI, 2000-318079/27.  
XX

Claim 7; Page 1476-1478; 1760pp; English.

Claim 7; Page 1476-1478; 1760pp; English

The present invention describes methods of obtaining immunogenic proteins from *Neisseria* genomic sequences. AAA81453 to AAA82414 represent specifically claimed *Neisseria meningitidis* genomic DNA sequences; AAA81260 to AAA81303 and AAB25620 to AAB25663 represent *Neisseria* DNA sequences and their corresponding proteins; AAA81254 to AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the isolation of *Neisseria meningitidis* DNA sequences; and AAA81322 to AAA81452 represent *Neisseria meningitidis* *WenB* polynucleotide ORF sequences, which are all used in the exemplification of the present invention. The nucleic acid sequences, protein sequences, and antibodies against them, can be used in the manufacture of a composition. The composition can be used as a medicament (or in the manufacture of a medicament) for treating, preventing or diagnosing infection due to *Neisseria* bacteria. For example, some of the identified proteins could be components of vaccines against *Meningococcus B* against all serotypes and/or against all pathogenic *Neisseriae*. Identification of sequences from the bacterium will also facilitate production of biological probes, particularly organism-specific probes. Attempts to make efficacious *Meningococcus B* vaccines have failed mainly due to antigen tolerance. Multivalent vaccines have also been tried but none have successfully overcome antigenic variability. The provision of further, complete sequences may provide an opportunity to identify secreted or surface exposed proteins that may be presumed targets for the immune system and which are not antigenically variable or at least more conserved than other more variable regions.

SQ Sequence 7919 BP; 2282 A; 1729 C; 1972 G; 1936 T; 0 other;

Alignment Scores:	
Pred. No.:	6.8e-111
Score:	1346.50
Percent Similarity:	82.9%
Best Local Similarity:	75.3%
Query Match:	5.2%
DB:	21
Gaps:	1
Length:	7919
Matches:	250
Conservative:	32
Mismatches:	47
Indels:	3
Gaps:	1

US-10-088-045-2 (1-502) x AAA81536 (1-7919)

Qy	56	GlutLeuProValIleAspAlaIleValThrHis:saIaProGluValaProProProValaAsp	77
Db	7917	GAACAGCCCGATTACGATGGCGTTACACACCACGCTCCGAACTGCTCTTGCAATGCAC	7855
Qy	78	ArgAspHis:spRoAlaIy:saValValValLySme:GluThrValGluLyValMetArgLeu	97
Db	7857	CGCGACTACCCCGCCAAAGTCGCGGTAAATAGAAACCGTCGAAAAAACCTATGCATG	7799
Qy	98	AlaAspGlyValGluTyrglnPheThrPheGlyGlyValaProGlyValaIleMetIle	117
Db	7797	GAAGACGGTGTGGAAATACCGCTACTGGACATTTGACGGCAGACTTCCGGGCCGTATGATC	7733
Qy	118	ArgValaArgGluGlyAspThrIleGluValaGlnPheSerAsnHisPheAspSerLySmeC	137
Db	7737	CGCGTACCCGAAAGCCGATACGATTGAAGTGAATTTTCCAAACAATCTTCTTCACCGTT	7678
Qy	138	ProHisAsenValaAspPheHis:saIaAlaIleThrGlyProGlyGlyValaGluValaSerPhe	157
Db	7677	CCGCACAAACGTGCATCTTCCACGGCGCTACCGGCGCAAGGCGGGCGGGCCGCACAACTTT	7618
Qy	158	ThrAlaProGlyHisThrSerThrPheSerPheLy:saIaLeuGlnProGlyLeuTyrrVal	177

Db	7617	ACCGCTCCGGGCGGACTTCCACATTACAGCTTCAAAAGCCCTGCAACCGGGCTGTGACATC	197
Qy	178	TYRHSQYSAIAVALAIPROVALGLYMETHSIIAIAASNGLYMETYRGLYLEU	197
Db	7557	TACCACTCGCGCTGCACCGGCTCGGTATGCACATCGCAACCGGTATACGGCTGAT	7498
Qy	198	LEUVALGIPROLYSGIUGLYLEUPROLYSVALASPLYSGLIUTRYTYVALMELGLI	217
Db	7497	TTGGTCAGACCTTAAGAAAGGCGCTGCCGAAAGTGATTAAGAGTTCTACATGCTCCAA	7438
Qy	218	ASPHERYTRHYSGLYLYSTYRGLYGLIUGIUGLYLEUGINPROCPHEASPMELIUDS	237
Db	7437	GACCTTCATACCAAGGCAAAAAAGGCGCGCAAGCTCTGCACACCTTGATATGACAA	7378
Qy	238	ALAILEARGLIUASPALGLIUTRYVALYALPHEASNGLYSERVALGLYALALEUTHGLY	257
Db	7377	GCCGTTGCCGCAACAGCTGAATACCGTGTATTCACACGCTACGTAAGGTGATCGCGGC	7318
Qy	258	GLIUSPNAIALUYSVALIYVALGLYGLIUTHRVALAUGLEUPHEVALGLYASNGLYGLY	277
Db	7317	GATTAACGCGCTGMAAGCCAAAGCAGGCGAACACTGATCGTATGACGTTGGTAAACGGCGCT	7258
Qy	278	PROASNEUTHRSERSEPHESISVALILEGLYGLIUPHEASPLYSVALIHPHEGLY	297
Db	7257	CCGAACCTTGATCTCTCTCCACGCTCATCGCGCAATCTTCGACAAAGTTATGTGAA	7198
Qy	298	GLYGLYLYSGLIUASNHIAASNILEGINTHRTHLEUILEPROLAGLYLYALAA	317
Db	7197	GCGCGCAACATGATTAACGAAACGTACAAAGCACATCGTCTCGCGCGGCTCGCC	7138
Qy	318	ILETHRGLUPHELYSVALASPVALPROGLYASPTRYVALLEUVALASPHIALAILEPHE	337
Db	7137	ATCGTCGAATTCAAAGTGCACATCCCGGCGACGTACACTTGGTGTGACCACTCATCTTC	7078
Qy	338	ARGVALPHEASNLVSGLYVALALEUGLYLEULEUYSVALIUGIUGLYGLIUSNHISGLY	357
Db	7077	CGCGCATTCACAAAGCGCGCACTGGGTCAATTGAAGTAAGGAGGCAAGAAACCTGAA	7018
Qy	358	ILETRYSERHIALYSGINTHRASPAVALATYRLEUPROGLIUGLYALAPROGLIHALAILE	377
Db	7017	ATCATGACTCAAAATTAAGTATGATACCGCTTACCGCGGTAAACGGTCACTCTGCT---	6961
Qy	378	ASPTHRGLINGLIUALPROLYSTHRPROALPROALA 389	
Db	6960	-----GCTTCGCTCCCGCAGCTTCTGCCCCGCA 6931	
RESULT 8			
AAL46493			
ID	AAL46493	standard; DNA; 691 BP.	
XX	AAL46493;		
XX	AC		
XX	AT	19-JUL-2002 (first entry)	
XX	DE	M catarrhalis MCA100170 gene SEQ ID NO: 1.	
XX	KV	Moraxella; vaccine; respiratory tract infection; antiinflammatory;	
XX	KW	auditory; antibacterial; otitis media; sinusitis; pneumonia; gene; ds.	
XX	OS	Moraxella catarrhalis.	
XX	PN	WO200218595-A2.	
XX	PD	07-MAR-2002.	
XX	PF	28-AUG-2001; 2001WO-CA01221.	
XX	PR	28-AUG-2000; 2000US-228294P.	
XX	PR	28-AUG-2000; 2000US-228295P.	
XX	PR	28-AUG-2000; 2000US-228296P.	
XX	PR	29-AUG-2000; 2000US-228438P.	
XX	PR	29-AUG-2000; 2000US-228439P.	

PR		29-AUG-2000;	2000US-2284410P.
PR		29-AUG-2000;	2000US-2284411P.
PR		29-AUG-2000;	2000US-2284412P.
PR		29-AUG-2000;	2000US-228443P.
PR		29-AUG-2000;	2000US-228511P.
PR		29-AUG-2000;	2000US-228512P.
PR		29-AUG-2000;	2000US-228742P.
PR		29-AUG-2000;	2000US-228773P.
PR		01-SEP-2000;	2000US-229465P.
PR		01-SEP-2000;	2000US-229474P.
PR		01-SEP-2000;	2000US-229475P.
PR		05-SEP-2000;	2000US-229478P.
PR		05-SEP-2000;	2000US-229740P.
PR		05-SEP-2000;	2000US-229803P.
PR		05-SEP-2000;	2000US-229804P.
PR		05-SEP-2000;	2000US-229805P.
PR		05-SEP-2000;	2000US-229806P.
PR		05-SEP-2000;	2000US-229809P.
PR		05-SEP-2000;	2000US-229811P.
PR		06-SEP-2000;	2000US-230214P.
PR		06-SEP-2000;	2000US-230250P.
PR		06-SEP-2000;	2000US-230252P.
XX			
PA	(AVET )	AVENTIS PASTEUR LTD.	
XX			
PI	Loosmore S,	Wang J,	Bradley B,
DR		Ochs M,	Yang Y;
XX			
DR	WPI; 2002-401721/43.		
P-PSDB:	AAOI7561.		
XX			
PT	Moraxella polypeptide and polymyxinectides useful as vaccine for immunizing a host e.g. humans against disease e.g. otitis media, pneumonia, caused by infection of the bacteria -		
XX			
PS	Claim 2; Fig 1; 277pp; English.		
XX			
CC	The present invention provides the protein and coding sequences of proteins from Moraxella catarrhalis. These can be used to produce vaccines which protect against M. catarrhalis infection, which can cause otitis media, respiratory infection, sinusitis, and pneumonia. The present sequence is a coding sequence of the invention.		
XX			
SQ	Sequence 691 BP; 184 A; 160 C; 161 G; 186 T; 0 other;		
Alignment Scores:			
Pred. No.:	4.09e-84	Length:	691
Score:	1034.00	Matches:	197
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	39.38%	Indels:	0
DB:	24	Gaps:	0
US-10-088-045-2 (1-502) x AAL46493 (1-691)			
OY	1	MeSerLySProThrIleuIlleysThrThrlleuIlleCySaIaleuSerIaleuMerLeu	20
Db	101	ATGTCTAAGCCTACTTGTATAAAAACAACCTTAATTGTCCTTAAGCATGATGCTC	160
OY	21	SeSGLYCVSSerSnglnAlaSprlySaIalagInPrOlysSerSerThrValAsPaLa	40
Db	161	AGGGTTTGAGCAATCACGGCAAGAAGCGGCCAACAAAAAGCAGACGGTAGACCT	220
OY	41	AlaAlLySThrAlaSnAlaSPasnaIalaSerglnglnIuhIsglnglygluleuPro	60
Db	221	GGCGGCAGACGAAGATGCAATATATCTGCTCTACAGAACATCAAGCGAGCTGCT	280
OY	61	VallIleSpaIalleValThrHisAlaProgluValPropOpRoVaIlaSPaTgASPhis	80
Db	281	GTCATTGATGCATTTGTTACGATCACCAAGATTCACACACCTGTTGACCCTGACAC	340
OY	81	ProAlaLySVaIVaIVaIIyMetLuthrVtAlGIuLySVaIMetArgLeuaIaaSPGLy	100
Db	341	CCCCCAAAGTGATGTAATAATGAAAACCGTGA AAAAGTCATGCGTCTGGCAGATGGC	400

QY 101 ValGIuTYrgInPheTrpThPheGIyGIyGlnValProGIyGlnMetIleArgValArg 120  
DB 401 GTGAAATATCAGTTTGGACATTTGGCGTCAAGTTCAGAGGCGAGATGATTCGTGTGCT 460  
QY 121 GIUGIYAAPTThIleGIuValGlnPheSerAsnHisProAspSerIleMetProHisAsn 140  
DB 461 GAAGGCGACACCATCGAAGTGAAGTTCMAACCCAGATTCMAAATCCCCCATATAT 520  
QY 141 ValAspPheHisAlaAlaThrGIyProGIyGIyGlnValAlaSerPheThrAlaPro 160  
DB 521 GTTACACTTTCACGCTGCACAGAGGCTTGGCGGCGGAGAACGCTCATTTACCGCACCG 580  
QY 161 GIYHSTHrSerThrPheSerPheIleAlaLeuGlnProGIyLeuTYrValTYrHisCys 180  
DB 581 GGTCAATACATCAACCTTATAGTTTAAAGCCTTACAGCCTGGTTGTATGCTATCACTGT 640  
QY 181 AlaValAlaProValAlaGIyMetHisIleAlaAsnGIyMetTYrGIyLeuIle 197  
DB 641 GCGGTTCCTCGTTGGCATGCACATTCCTAATGGCATGATGATGTTGATT 691

RESULT 9  
AAS59518  
ID AAS59518 standard; DNA; 37337 BP.  
XX AAS59518;  
AC AAS59518;  
DT 13-FEB-2002 (first entry)  
XX  
DE Propionibacterium acnes immunogenic protein encoding DNA #13.  
XX  
KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;  
KM uveitis; endophthalmitis; bone joint; central nervous system; ELISA;  
NM inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
KM dermatological; osteopathic; neuroprotectant; ds.  
OS Propionibacterium acnes.  
XX  
PN WO200181581-A2.  
XX  
PD 01-NOV-2001.  
XX  
PF 20-APR-2001; 2001WO-US12865.  
XX  
PR 21-APR-2000; 2000US-199047P.  
PR 02-JUN-2000; 2000US-208841P.  
PR 07-JUL-2000; 2000US-216747P.  
XX  
PA (CORI-) CORIXA CORP.  
XX  
PI Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;  
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;  
XX  
DR WPI; 2001-616774/71.  
XX  
PT Propionibacterium acnes polypeptides and nucleic acids useful for  
PT vaccinating against and diagnosing infections, especially useful for  
PT treating acne vulgaris -  
XX  
PS Claim 1; SEQ ID NO 13; 1069pp; English.  
XX  
CC Sequences AAS59506-AAS59804 represent DNA molecules encoding  
CC Propionibacterium acnes immunogenic polypeptides. The proteins and their  
CC associated DNA sequences are used in the treatment, prevention and  
CC diagnosis of medical conditions caused by P. acnes. The disorders include  
CC SAPHO syndrome (synovitis, acne, pustulosis, hypertosis and  
CC osteomyelitis), uveitis and endophthalmitis. P. acnes is also involved  
CC in infections of bone, joints and the central nervous system, however it  
CC is particularly involved in the inflammatory lesions associated with acne  
CC vulgaris. A method for detecting the presence or absence of P. acnes in a  
CC patient comprises contacting a sample with a binding agent that binds to  
CC the proteins of the invention and determining the amount of bound protein  
CC in the sample. The polypeptides may be used as antigens in the production

CC of antibodies specific for P. acnes proteins. These antibodies can be  
CC used to downregulate expression and activity of P. acnes polypeptides and  
CC therefore treat P. acnes infections. The antibodies may also be used as  
CC diagnostic agents for determining P. acnes presence, for example, by  
CC enzyme linked immunosorbent assay (ELISA). This sequence encodes the  
CC polypeptides shown in AAU42549-42961, AAU67480 and AAU67481.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 37337 BP; 7213 A; 12017 C; 11261 G; 6843 T; 3 other;

Alignment Scores:  
Pred. No.: 5.64e-20 Length: 37337  
Score: 349.00 Matches: 92  
Percent Similarity: 47.10% Conservative: 46  
Best Local Similarity: 31.40% Mismatches: 121  
Query Match: 13.29% Indels: 34  
DB: 23 Gaps: 10

US-10-088-045-2 (1-502) x AAS59518 (1-37337)

QY 72 ValProProProVal-----AspArgAspHisProAlaIleValValIleMetGlu 89  
DB 17911 GTGCTACACACCTGATGACCGGACGCTTACCGGATGACGCTACCGCCAG----- 17964  
QY 90 ThrValGluIleValMetArgIleAlaAspGluValGluTYrGlnIleThrPheGly 109  
DB 17965 -----GAGTCGGTGGACGATGCGCCCGAAACCAACCATCGACCGGATGACTACAC 18018  
QY 110 GIYGIuValProGIyGlnMetIleArgValArgGluGIyAspThrIleGluValGlnPhe 129  
DB 18019 GGCCTGTACATGCGTCCGGTATATCCAGCGGATGCGGACGAGATGAGGTCATCTG 18078  
QY 130 SerAsnHisProAspSerIleMetProHisAsnValAspPheHisAlaIleThrGlyPro 149  
DB 18079 GTCACACAG-----GGCACTAATGAGGTCACTTGTGACTTTCACCGCGTACT----- 18126  
QY 150 GIYGIyGIyAlaGluAlaSerPheThr-----AlaProGIyHisThr 163  
DB 18127 -----GTCTCCCGCCACGAGGCTATCGCACCATCGCTCGGGGACAGAG 18171  
QY 164 SerThrPheSerPheIleAlaLeuGlnProGIyLeuTYrValTYrHisCysAlaValAla 183  
DB 18172 TTGACTACACATTCACCGCTGATCGCGCGGATGCGGTCACTTGTCTCCACCGCG 18231  
QY 184 ProValGIyMetHisIleAlaAsnGIyMetTYrGIyLeuIleValGluProIleGlu 203  
DB 18232 CCAATGTCCGCCCATATTCAGACGAGCATGTTCCGTGCGTCACTGCGCCGACAGAC 18291  
QY 204 GIYLeuProIleValAspIleGluTYrTYrValMetGlnGIyAspPheTYrThrIleGly 223  
DB 18292 ---CTGCTCGGCTATCGAGGAGTTTACTGTGCTGCGAGGAGCTTACCTCACTGAG 18348  
QY 224 LysTYrGIyGlnGlnGIyLeuGlnProPheAspMetGluIleAlaIleArgGluAspAla 243  
DB 18349 CACAATGGGGCGGAA-----GTTAATATCGCAAGATTGCCAAGACAGACCCCT 18396  
QY 244 GIUTYrValValPheAsnGIySerValGIyAlaLeuThrGIyGluAsnAlaLeuIleVal 263  
DB 18397 GATCTGACGATGTTTCAACGCTCACGCCAACCAAGTACCTTTGAG---CCCTCAAAAGCC 18453  
QY 264 LysValGIyGluThrValArgLeuPheValGIyAsnGIyGluProAsnLeuThrSerSer 283  
DB 18454 CCGGTGGTGGAGGAGTGGAGTCTGGGTGCTGCGCGCTGTCTCAAGCGGGGCTTCTCC 18513  
QY 284 PheHisValIleGIyGluIlePheAspIleValHisPheGlnGIy----- 298  
DB 18514 TTTCACTGTGTGGGAACCAAGTTTGCACACTGTTTAAAGAGGGGCCCTACACCTTAA 18573  
QY 299 ---GIYLysGIyGluAsnHisAsnIleGlnThrIleThrIleProAlaGIyAlaAla 317  
DB 18574 CGTGGCAATCCGGAAGTGGCGGCTGCAAGCCCTGCACTGCGCACGCCCGAGGGCGGT 18633



QY 318 ILeThrGluPheLysValAspValProGlyAspTyrValLeuValAspHisAlaIlePhe 337  
Db 18634 TTGTGCGAGATGCTCTTGTAGGAGCGGAGCGCTACACCTTTGTAACATTCC--TTT 18690  
QY 338 ArgAlaPheAsnLysGlyAlaLeuGlyTyrLeuLysVal 350  
Db 18691 GTGAGATGAGCGCTGAGCTTAAAGATTCATTGAGGTG 18729  
RESULT 10  
AAZ53036  
ID AAZ53036 standard; DNA; 399 BP.  
AC AAZ53036;  
XX  
XX  
XX  
XX 21-MAR-2000 (first entry)  
DE Neisseria gonorrhoeae ORF 007 partial DNA sequence SEQ ID NO:43.  
XX  
XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;  
KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;  
KM antibacterial; gene therapy; ds.  
XX  
XX  
OS Neisseria gonorrhoeae.  
XX  
XX  
XX MO9957280-A2.  
PN 11-NOV-1999.  
PD 30-APR-1999; 99MO-US09346.  
XX  
XX 01-MAY-1998; 98US-0083758.  
XX 31-JUL-1998; 98US-0094869.  
XX 02-SEP-1998; 98US-0098994.  
XX 02-SEP-1998; 98US-0099062.  
XX 09-OCT-1998; 98US-0103749.  
XX 09-OCT-1998; 98US-0103794.  
XX 09-OCT-1998; 98US-0103796.  
XX 25-FEB-1999; 99US-0121528.  
XX  
XX  
XX (CHIR) CHIRON CORP.  
XX (GENO-) INST GENOMIC RES.  
XX  
XX Frazer C, Galeotti C, Grandi G, Hickey E, Maignani V, Mora M;  
PI Petersen J, Piza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;  
PI Tettelin H, Venter JC;  
XX  
XX WPI; 2000-062150/05.  
XX P-PSDB; AAY74274.  
XX  
XX Novel Neisserial polypeptides predicted to be useful antigens for  
PT vaccines and diagnostics -  
XX  
XX  
XX Claim 7; Page 180; 1453pp; English.  
XX  
XX AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941  
XX represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides  
XX and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent  
XX PCR primers used in the exemplification of the present invention. The  
XX polypeptides, the polynucleotides, antibodies and compositions of  
XX the invention can be used as vaccines, as diagnostic reagents, and as  
XX immunogenic compositions. The polypeptides can be used in the  
XX manufacture of medicaments for treating or preventing infection due to  
XX Neisserial bacteria (e.g. meningitis and septicemia), to detect the  
XX presence of Neisseria bacteria, or to raise antibodies. They may also  
XX be used to screen for agonists or antagonists, which may themselves  
XX have use as antibacterial agents. The polynucleotides of the invention  
XX may also be used in gene therapy protocols.  
XX  
XX Sequence 399 BP; 118 A; 122 C; 93 G; 66 T; 0 other;  
XX  
XX  
XX Alignment Scores: 1.64e-12 Length: 399  
XX  
XX Pred. No.:

Score: 237.00 Matches: 50  
Percent Similarity: 60.58% Conservative: 13  
Best Local Similarity: 48.08% Mismatches: 19  
Query Match: 9.03% Indels: 2  
DB: 21 Gaps: 2  
US-10-088-045-2 (1-502) x AAZ53036 (1-399)  
QY 398 GILYVAlaThrTyrAspSerAsnCysAlaAlaCysHisGlnProAspGlyLysGlyVal 417  
Db 88 GGGCAAAAAGTGTAGCATTCATCTCATCGCTGCCACGGCAAGAAAGGGAAGGGCC 147  
QY 418 ProAsnAlaPheProPheLeuAlaAsnSerAspTyr--LeuAsnAlaAspHisAlaArg 436  
Db 148 GGCACGCGCTTCCCTCCGCTTTCCGCTGCACTATATATGAACAACCGCACGCTCG 207  
QY 437 AlaAlaSerTyrValAlaAsnGlyLeuSerGlyLysIleThrValAsnGlyAsnGlnTyr 456  
Db 208 CTGCACAGCATGGTC--AAAGCATCAACGGTACATCAAGTCAACGGCAAAACCTAC 264  
QY 457 GluSerValMetProAlaIleAlaLeuSerAspGlnGlnIleAlaAsnValIleThrTyr 476  
Db 265 AACGATTGATGCGCCGCAACCGCATCAGCATGCGGACATTGCCCGCTCGCACTTAT 324  
QY 477 ThrLeuAsnSerPheGlyAsnLysGlyGlnLeuSerAlaAspValAlaLysAla 496  
Db 325 ATCATGACACGCTTGTGACAAACGGCGGAGCGTTACCGAAAGACGTAAACAGGCA 384  
QY 497 LysLysThrLys 500  
Db 385 AAAGCCAAAAA 396  
RESULT 11  
ABZ39100  
ID ABZ39100 standard; DNA; 552 BP.  
XX  
XX ABZ39100;  
AC  
XX  
XX 07-MAR-2003 (first entry)  
DT  
XX  
XX N. gonorrhoeae nucleotide sequence SEQ ID 2789.  
DE  
XX  
XX Antibacterial; infection; vaccine; gene therapy; gene; ds.  
KW  
XX  
XX Neisseria gonorrhoeae.  
OS  
XX  
XX WO200279243-A2.  
XX  
XX 10-OCT-2002.  
XX  
XX 12-FEB-2002; 2002WO-1B02069.  
XX  
XX 12-FEB-2001; 2001GB-0003424.  
XX  
XX (CHIR-) CHIRON SPA.  
XX  
XX Fontana MR, Piza M, Maignani V, Monaci E;  
PI WPI; 2003-058415/05.  
XX P-PSDB; ABP78130.  
XX  
XX New protein from Neisseria gonorrhoeae, useful for the manufacture of a  
PT medicament for treating or preventing N. gonorrhoeae infection -  
XX  
XX Disclosure; Page 392; 815pp; English.  
XX  
XX The present invention relates to proteins from Neisseria gonorrhoeae.  
XX Also disclosed are the nucleic acid molecules encoding the proteins and  
XX antibodies that specifically bind to the proteins. The composition  
XX comprising the protein, nucleic acid or antibody is useful for the  
XX manufacture of a medicament for treating or preventing N. gonorrhoeae  
XX infection, this may be in the form of a vaccine or gene therapy.  
XX Sequences given in records ABZ37706-ABZ42016 represent nucleic acid

CC molecules of the invention.

SQ Sequence 552 BP; 175 A; 167 C; 116 G; 94 T; 0 other;

Alignment Scores:

Pred. No.:	2,536-12	Length:	552
Score:	237.00	Matches:	50
Percent Similarity:	60.56%	Conservative:	13
Best Local Similarity:	48.08%	Mismatches:	39
Query Match:	9.03%	Indels:	2
	25	Gaps:	2

US-10-088-045-2 (1-502) x AB239100 (1-552)

```

QY 398 GGLYSAIATHTYRASPSEASANCYSAIAAIAcYSHISGInProAspGlyLYSGlyVal 417
    |||:::|
DB 175 GGGCAAAAAGTACGAATCCACATCGCTCCACGGCAAGAAAGGGAGGCGC 234
QY 418 ProAsnAlaPheProProLeuAlaAsnSerAspTYR--LeuAsnAlaAspHisAlaArg 436
    |||:::|
DB 235 GGCACCTGGCTTCTCCGCTTTCGGCTGACATATATATGAACAACCGACGCTCTG 294
QY 437 AlaAlaSerIleValAlaAsnGlyLeuSerGlyLYSleThrValAsnGlyAsnGlnTYR 456
    |||:::|
DB 295 CTGCACACATGCTC--AAAGCATCAACGCTACAAATCAAGTCAACGGCAAAACCTAC 351
QY 457 GluSerValMetProAlaIleAlaLeuSerAspGlnGlnIleAlaAsnValIleThrTYR 476
    |||:::|
DB 352 AACGATTCATGCCCGCAACCCCATCCAGCATGGACATTTGCCCGCTGCCCATTTAT 411
QY 477 ThrLeuAsnSerPheGlyAsnLYSGlyGlnLeuSerAlaAspAspValAlaLYSA 496
    |||:::|
DB 412 ATCATGAACGCTTTGACACACGGCGCGGAGCTTACCGAAAGAGCTAAACAGGCA 471
QY 497 LysIleThrLYS 500
    |||
DB 472 AAAGCAAAAA 483

RESULT 12
AAZ53038
ID AAZ53038 standard; DNA; 396 BP.
XX
AC AAZ53038;
XX
DT 21-MAR-2000 (first entry)
XX
DE Neisseria meningitidis ORF 007 partial DNA sequence SEQ ID NO:47.
XX
KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
KW antibacterial; gene therapy; ds.
XX
OS Neisseria meningitidis.
XX
PN WO957280-A2.
XX
PD 11-NOV-1999.
XX
PF 30-APR-1999; 99WO-US09346.
XX
PR 01-MAY-1998; 98US-0083758.
PR 31-JUL-1998; 98US-0094869.
PR 02-SEP-1998; 98US-0098994.
PR 02-SEP-1998; 98US-0099062.
PR 09-OCT-1998; 98US-0103749.
PR 09-OCT-1998; 98US-0103794.
PR 09-OCT-1998; 98US-0103796.
PR 25-FEB-1999; 99US-0121528.
XX
PA (CHIR ) CHIRON CORP.
PA (GENO-) INST GENOMIC RES.
XX
PI Fraser C, Galeotti C, Grandi G, Hickey E, Massignani V, Mora M;

```

PI Petersen J, Piza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;  
PI Tectelin H, Venter JC;  
XX  
XX WPI: 2000-062150/05.  
DR P-PSDB; AAY74276.  
XX  
XX Novel Neisserial polypeptides predicted to be useful antigens for  
PT vaccines and diagnostics  
PS Claim 7; Page 180-181; 1453pp; English.

XX AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941  
CC represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides  
CC and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ5473 represent  
CC PCR primers used in the exemplification of the present invention. The  
CC polypeptides, the polynucleotides, antibodies and compositions of  
CC the invention can be used as vaccines, as diagnostic reagents, and as  
CC immunogenic compositions. The polypeptides can be used in the  
CC manufacture of medicaments for treating or preventing infection due to  
CC presence of Neisseria bacteria, or to raise antibodies. They may also  
CC be used to screen for agonists or antagonists, which may themselves  
CC have use as antibacterial agents. The polynucleotides of the invention  
CC may also be used in gene therapy protocols.

SQ Sequence 396 BP; 118 A; 125 C; 92 G; 61 T; 0 other;

Alignment Scores:

Pred. No.:	3,666-11	Length:	396
Score:	222.00	Matches:	44
Percent Similarity:	56.31%	Conservative:	14
Best Local Similarity:	42.72%	Mismatches:	45
Query Match:	8.45%	Indels:	0
	21	Gaps:	0

US-10-088-045-2 (1-502) x AAZ53038 (1-396)

```

QY 398 GGLYSAIATHTYRASPSEASANCYSAIAAIAcYSHISGInProAspGlyLYSGlyVal 417
    |||:::|
DB 88 GGGCAAAAAGTACGAATCCACATCGCTCCACGGCAAGAAAGGGAGGCGC 147
QY 418 ProAsnAlaPheProProLeuAlaAsnSerAspTYRLeuAsnAlaAspHisAlaArgAla 437
    |||:::|
DB 148 GGAACCATGTTTCCGCGCTCTACCGCTCCGATCATCATGAAGAAACCGACAGTGTG 207
QY 438 AlaSerIleValAlaAsnGlyLeuSerGlyLYSleThrValAsnGlyAsnGlnTYR 457
    |||:::|
DB 208 CTGCACACATGCTTCAAGGATCAACGATCAAGTCAACGCAAAACCTACAC 267
QY 458 SerValMetProAlaIleAlaLeuSerAspGlnGlnIleAlaAsnValIleThrTYR 477
    |||:::|
DB 268 GGAATTCATGCCCGCCACATCGCATACGATGCGAGCATTCGCCCGCTGCCACTTATAC 327
QY 478 LeuAsnSerPheGlyAsnLYSGlyGlnLeuSerAlaAspAspValAlaLYSA 497
    |||:::|
DB 328 ATGAACGCTTTGACACACGGCGGAGAGCTTACCGAAAGAGCTAAACAGGCAAAA 387
QY 498 LysIleThrLYS 500
    |||
DB 388 AACCAAAAA 396

RESULT 13
AAZ53037
ID AAZ53037 standard; DNA; 402 BP.
XX
AC AAZ53037;
XX
DT 21-MAR-2000 (first entry)
XX
DE Neisseria meningitidis ORF 007 partial DNA sequence SEQ ID NO:45.
XX
KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;

```

KM antibacterial; gene therapy; ds.  
 XX Neisseria meningitidis.  
 OS  
 XX MO9957280-A2.  
 PN  
 XX 11-NOV-1999.  
 PD  
 XX 30-APR-1999; 99WO-US09346.  
 PF  
 XX 01-MAY-1998; 98US-0083758.  
 PR 31-JUL-1998; 98US-0094869.  
 PR 02-SEP-1998; 98US-0098994.  
 PR 02-SEP-1998; 98US-0099062.  
 PR 04-OCT-1998; 98US-0103749.  
 PR 09-OCT-1998; 98US-0103794.  
 PR 09-OCT-1998; 98US-0103796.  
 PR 25-FEB-1999; 99US-0121528.  
 XX  
 XX (CHIR) CHIRON CORP.  
 PA (GENO-) INST GENOMIC RES.  
 XX  
 XX Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;  
 PI Petersen J, Pizze M, Rappuoli R, Ratti G, Scalato E, Scarcellì M;  
 PI Tettelin H, Venter JC;  
 DR WPI: 2000-062150/05.  
 XX P-PSDB; AAY74275.  
 DR  
 PT Novel Neisserial polypeptides predicted to be useful antigens for  
 PT vaccines and diagnostics  
 PS  
 XX Claim 7; Page 180; 1453pp; English.  
 XX  
 CC AA253015 to AA254536, AA254577 to AA254615, and AAY74253 to AAY75941  
 CC represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides  
 CC and polypeptides. AA254537 to AA254576 and AA254616 to AA255473 represent  
 CC PCR primers used in the exemplification of the present invention. The  
 CC polypeptides, the polynucleotides, antibodies and compositions of  
 CC the invention can be used as vaccines, as diagnostic reagents, and as  
 CC immunogenic compositions. The polypeptides can be used in the  
 CC manufacture of medicaments for treating or preventing infection due to  
 CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the  
 CC presence of Neisseria bacteria, or to raise antibodies. They may also  
 CC be used to screen for agonists or antagonists, which may themselves  
 CC have use as antibacterial agents. The polynucleotides of the invention  
 CC may also be used in gene therapy protocols.  
 CC  
 XX  
 XX Sequence 402 BP; 122 A; 125 C; 93 G; 62 T; 0 other;  
 SQ  
 Alignment Scores:  
 Pred. No.: 3, 73e-11 Length: 402  
 Score: 222.00 Matches: 44  
 Percent Similarity: 56.31% Conservative: 14  
 Best Local Similarity: 42.72% Mismatches: 45  
 Query Match: 8.45% Indels: 0  
 DB: 21 Gaps: 0  
 US-10-088-045-2 (1-502) x AA253017 (1-402)  
 QY 398 Gtlysalathtyrspseranrcyvalaalacyahieglnprorasglytylsglyval 417  
 DB 88 GGGCAAAAGTGTACCAATCCAACTCGCTGCGCAAGGCAAAAGGCGCAAGCCCGC 147  
 QY 418 ProaenallaheproProleuallaenSeraspTYrleuasnallaasphisalaargala 437  
 DB 148 GGAAACCATGTTCCCGCCCTTACCGCTCGACTTCATCATGAAAAAACCGCAGGTCTG 207  
 QY 436 AlaserllevalaalsnglyleuserglylyllethThValasnglylsngintyrglu 457  
 DB 208 CTGCACGACATGTCAAAGGACATCAAGGACATCAATCAAGGCAAGGCAAGGCAAGC 267  
 QY 458 SerValmetProAlaileAlaaleuSerAspGlnGlnleAlaasnValilethThrTYrThr 477

DB 268 GGATTCATGCGCCGCAACCGCCATTCAGCGATGCGGACATTCGCCCGCCCTATATC 327  
 QY 478 LeuAenSerPheGlyYasnlysglylgluSerAlaAspAspValAlaYsaAlaYs 497  
 DB 328 ATGAACGCTTTCACACACGCGCGCGAGCGTTACCGAAAAAGACGTAAACAGCGCAAA 387  
 QY 498 LysThrLys 500  
 DB 388 AGCAAAAAA 396  
 RESULT 14  
 ID AAA81733/C  
 ID AAA81733 standard; DNA; 12893 BP.  
 AC AAA81733;  
 DT 04-DEC-2000 (first entry)  
 XX  
 XX N. meningitidis partial DNA sequence gnm\_280 SEQ ID NO:280.  
 DE  
 XX Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;  
 KW antigen; vaccine; diagnosis; infection; antibacterial; identification;  
 KW Meningococcus B; MenB; ds.  
 XX  
 XX Neisseria meningitidis.  
 OS  
 XX MO20002430-A2.  
 XX  
 XX 20-APR-2000.  
 PD  
 XX 08-OCT-1999; 99WO-US23573.  
 PF  
 XX 09-OCT-1998; 98US-0103794.  
 PR 30-APR-1999; 99US-0132068.  
 XX  
 XX (CHIR) CHIRON CORP.  
 PI Fraser CM, Hickey E, Petersen J, Tettelin H, Venter JC;  
 PI Masignani V, Galeotti C, Mora M, Ratti G, Scarcellì M, Scalato V;  
 PI Rappuoli R, Pizze M;  
 DR WPI: 2000-318079/27.  
 XX  
 PT Isolated nucleotide sequences of Neisseria meningitidis which can be  
 PT used in the diagnosis and treatment of N. meningitidis infection and  
 PT other Neisserial infections, for example, N. gonorrhoea  
 PS  
 XX Claim 7; Page 1561-1565; 1760pp; English.  
 XX  
 CC The present invention describes methods of obtaining immunogenic  
 CC proteins from Neisseria genomic sequences. AAA81453 to AAA82414  
 CC represent specifically claimed Neisseria meningitidis genomic DNA  
 CC sequences; AAA81560 to AAA81303 and AAB25620 to AAB25663 represent  
 CC Neisseria DNA sequences and their corresponding proteins; AAA81254 to  
 CC AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the  
 CC isolation of Neisseria meningitidis DNA sequences; and AAA81322 to  
 CC AAA81452 represent Neisseria meningitidis MenB polynucleotide ORF  
 CC sequences, which are all used in the exemplification of the present  
 CC invention. The nucleic acid sequences, protein sequences, and antibodies  
 CC against them, can be used in the manufacture of a composition. The  
 CC composition can be used as a medicament (or in the manufacture of a  
 CC medicament) for treating, preventing or diagnosing infection due to  
 CC Neisserial bacteria. For example, some of the identified proteins could  
 CC be components of vaccines against Meningococcus B; against all serotypes;  
 CC and/or against all pathogenic Neisseriae. Identification of sequences  
 CC from the bacterium will also facilitate production of biological probes,  
 CC particularly organism-specific probes. Attempts to make efficacious  
 CC Meningococcus B vaccines have failed mainly due to antigen tolerance.  
 CC Multivalent vaccines have also been tried but none have successfully  
 CC overcome antigenic variability. The provision of further, complete  
 CC sequences may provide an opportunity to identify secreted or surface  
 CC exposed proteins that may be presumed targets for the immune system and

CC which are not antigenically variable or at least more conserved than  
 CC other more variable regions.

XX  
 SQ Sequence 12893 BP; 2829 A; 3164 C; 3741 G; 3159 T; 0 other;

# Alignment Scores:

Pred. No.:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
Score:	3,91e-09	12893	64	32	84	42
Percent Similarity:	222.00					
Best Local Similarity:	43.44%					
Query Match:	28.96%					
DB:	21					6

US-10-088-045-2 (1-502) x AAA81733 (1-12893)

```

QY 301 G|YGLUASNHASNI|EGINThrThleuileProalaGlyAlaIleThrGlu 320
DB 10503 GGCACACACATCAATGACAA-----CCTTATAGTGATTAATTAAACGAG 10456
QY 321 Phe|yVal|Asp|Val|ProGly|Asp|Tyr----- 329
DB 10455 TACGGCTTCGCTCGCTTG--CCGTACTATTGTACTGTCTGCGGCTTCGCTTATC 10397
QY 330 Val|euVal|Asp|Ile|Ala|Ile|Phe|Arg|Ala|Phe|Asn|yS|G|Ala|Leu|G|Ile|Leu|yS 349
DB 10396 CTGATTTTGTATTAATCACTACTACTGATTCGGCAAAAATCTTCAACCGCTACGC 10337
QY 350 Val|G|u|G|I|G|u|G|u|A|N|H|S|G|I|U|I|E|T|S|E|N|H|I|S|G|I|N|T|H|-----Asp|Ala|Val 367
DB 10336 -----AAATCAACACACAAAGTTTCAAGTAAACATGAAACACATGACACACC 10286
QY 368 Tyr|Leu|ProGlu-----G|y|Ala|ProGlu|Ala|Ile|Asp|Thr 379
DB 10285 CGACTGCGACCGCCCTCGCTTGGGCTGCTTGGCCCGCTTGGCCGCGACAC 10226
QY 380 G|N|G|u|A|I|A|P|ro|y|Th|P|ro|Ala|P|ro|Ala|A|S|n|L|e|u|G|I|N|I|E|y|S|A|G|I|y|S 399
DB 10225 AGCATCATGACAA-----GGGCAA 10205
QY 400 Al|a|Th|r|Tyr|Asp|Ser|Arg|Ala|Ile|Ala|C|y|H|I|S|G|I|N|P|ro|Asp|G|I|y|S|G|I|y|Ala|Pro|Asn 419
DB 10204 AAAGTATGACAACTCACTCGCTGCTGCGCAAAAAGGCGCAAGCCGCGAACC 10145
QY 420 Al|a|P|he|P|ro|P|ro|Leu|Ala|Asn|Ser|Asp|Tyr|Leu|Asn|Ala|Asp|Ile|Ala|Ile|Ala|Ser 439
DB 10144 ATGTTTCCGCGCTTACCGCTCGCATCATCTGAAAAACCGCGAGTCTGCTGAC 10085
QY 440 I|l|e|V|a|l|A|a|S|n|G|I|y|S|E|r|G|I|y|S|I|E|T|H|V|a|l|S|n|G|I|y|S|n|G|I|y|S|E|r|V|a|l 459
DB 10084 AGCATGTCAAAGGATCAACGGTACATCAAGTCAACGGCAAAACCTACACGATTC 10025
QY 460 Met|P|ro|Ala|Ile|Ala|Leu|Ser|Asp|G|I|N|I|E|A|A|S|n|V|a|l|I|E|T|H|Tyr|Thr|Leu|Asn 479
DB 10024 ATGCCCGCAACCGCATCAGCGATGCGGACATTCGCCCGCTCGCATTTATCATTCAC 9965
QY 480 Ser|P|he|G|I|y|Asn|yS|G|I|y|G|I|N|Leu|Ser|Ala|Asp|Ser|Val|Ala|yS|Ala|yS|Thr 499
DB 9964 GCTTTTACAAACGGCGGAGAGCTTACCCAAAAAGACTTAAACAGGCAAAAGCAA 9905
QY 500 Lys 500
DB 9904 AAA 9902

```

## RESULT 15

AAFP21608/c standard; DNA; 349980 BP.

```

ID AAF21608;
AC AAF21608;
XX
XX
DT 13-MAR-2001 (first entry)
XX
XX Neisseria meningitidis B nucleotide sequence SEQ ID NO:109.
XX

```

KM Neisseria meningitidis; Neisseria gonorrhoeae; immunogenic; vaccine;  
 KM diagnosis; antigen; detection; infection; gene therapy; antibacterial;  
 KM ds.

OS Neisseria meningitidis.

PN WO200066791-A1.

PD 09-NOV-2000.

PF 08-MAR-2000; 2000MO-US05928.

PR 30-APR-1999; 99US-0132068.

PR 08-OCT-1999; 99MO-US23573.

PR 28-FEB-2000; 2000GB-0004695.

PA (CHIR) CHIRON CORP.

PA (GENO-) INSTR GENOMIC RES.

PI Pizza M, Hickey E, Peterson J, Tetrelin H, Venter JC, Masignani V;  
 PI Galeotti C, Mora M, Ratti G, Scarcelli M, Scariato V, Rappuoli R;  
 PI Frazer CM, Grandi G;

XX MPI; 2000-647603/62.

XX Neisseria meningitidis B full length genome sequence and open reading  
 PT frames are used to detect, treat and prevent Neisserial infections -

XX Claim 7; Appendix A; 692pp; English.

CC The present invention describes the full length genome of  
 CC Neisseria meningitidis B (NMB). The sequences in AAF21544 and AAF21607  
 CC to AAF21613 represent fragments of the NMB genomic sequence, as the  
 CC sequence was too long to go in a record on its own it was split into 8  
 CC sequences which overlap each other at the beginning and end of each  
 CC sequence by 49980 bp (i.e. the last 49980 bp of AAF21544 is repeated at  
 CC the beginning of AAF21607, the last 49980 bp of AAF21607 are repeated at  
 CC the beginning of AAF21608, and so on). AAF21545 to AAF21588 encode the  
 CC Neisseria proteins given in AAB58550 to AAB58593, and AAF21589 to  
 CC AAF21606 represent PCR primers which are used in the exemplification of  
 CC the present invention. The NMB genome and fragments from it have  
 CC antibacterial activity, and can be used in vaccines and gene therapy.  
 CC Neisseria nucleic acids, proteins and/or antibodies which binds to the  
 CC proteins can be used in compositions for treating or preventing infection  
 CC due to Neisserial bacteria or as a diagnostic reagent for detecting the  
 CC presence of Neisserial bacteria or of antibodies raised to Neisserial  
 CC bacteria. Computers, computer memory, computer storage medium or computer  
 CC databases can be used in a search to identify open reading frames (ORFs)  
 CC or coding sequences within the NMB genome. The DNA sequences provide  
 CC further opportunities to find antigenic or immunogenic proteins which are  
 CC more effective in vaccines than the outer membrane proteins currently  
 CC used.

XX SQ Sequence 349980 BP; 82523 A; 82940 C; 96712 G; 87805 T; 0 other;

## Alignment Scores:

Pred. No.:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
Score:	3,28e-07	349980	64	32	84	42
Percent Similarity:	222.00					
Best Local Similarity:	43.44%					
Query Match:	28.96%					
DB:	21					6

US-10-088-045-2 (1-502) x AAF21608 (1-349980)

```

QY 301 G|YGLUASNHASNI|EGINThrThleuileProalaGlyAlaIleThrGlu 320
DB 148400 GGCACACACATCAATGACAA-----CCTTATAGTGATTAATTAAACGAG 148353
QY 321 Phe|yVal|Asp|Val|ProGly|Asp|Tyr----- 329
DB 148352 TACGGCTTCGCTCGCTTG--CCGTACTATTGTACTGTCTGCGGCTTCGCTTATC 148294

```

```
QY 330 ValIeuValAspHisAlaIlePheAsnIleGlyAlaLeuGlyIleLeuLys 349
DB 148293 CTGATTTTGTATTCACATCATCTACTGATTCGGCGGCAAAATCTTTCACCGCGTACGC 148234
QY 350 ValIGluGlyGluGlnAsnHisGluIleTyrSerHisIleGlnThr-----AspAlaVal 367
DB 148233 -----AAATTCACAACAACAGGTTTCAACATGAAAGAAACCAATGAACACACACC 148183
QY 368 TyrLeuProGlu-----GlyAlaProGlnAlaIleAspThr 379
DB 148182 CGACTGCCGACCGCGCTGCTTGGGCTGCTTGGCGCGCGCGCTTCTGCGCGGACAC 148123
QY 380 GlnGluAlaProLysThrProAlaProAlaAsnLeuGlnGluGlnIleLysAlaGlyLys 399
DB 148122 AGCATCATGACAA-----GGGCAA 148102
QY 400 AlaThrTyrAspSerAsnGlyAlaIleCysHisGlnProAspGlyLysGlyValProAsn 419
DB 148101 AAAGTGTACGAATCCACTGCTGCTGCGCGCAAAAGGCGAAAGCGCGGACACC 148042
QY 420 AlaPheProProLeuAlaAsnSerAspTyrLeuAsnAlaAspHisAlaArgAlaIleAsn 439
DB 148041 ATTTTCCGCGCTTACCGCTCGACTCATCATGAAAAAACCGACGCTGCTGCAC 147982
QY 440 IleValAlaAsnGlyLeuSerGlyLysIleThrValAsnGlyAsnGlnTyrGlnSerVal 459
DB 147981 AGCATGTATCAAGGATCAACGCGTACATCAAGTCAACGCGCAAAACCTTCAACGATTC 147922
QY 460 MetProAlaIleAlaLeuSerAspGlnGlnIleAlaAsnValIleThrTyrThrLeuAsn 479
DB 147921 ATGCCCGCAACCGCATCAACGATGCGGACATTCGCGCGCGCTTATATCATGAAC 147862
QY 480 SerPheGlyAsnIleGlyGlnLeuSerAlaAspAspValAlaLysAlaIleThr 499
DB 147861 GCCTTGACACACGCGCGGACCGCTTACCGAAAAAGACTTAAACAGGCAAAAGACAA 147802
QY 500 Lys 500
DB 147801 AAA 147799
RESULT 16
ID AAA81490 standard; DNA; 1437668 BP.
XX
AC AAA81490;
XX
DT 04-DEC-2000 (first entry)
XX
DE N. meningitidis B full length genome DNA sequence SEQ ID NO:1068.
XX
KW Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;
KM antigen; vaccine; diagnosis; infection; antibacterial; identification;
KM Meningococcus B; MenB; de.
XX
OS Neisseria meningitidis.
XX
PN MO00022430-A2.
XX
PD 20-APR-2000.
XX
PF 08-OCT-1999; 99WO-US23573.
XX
PR 09-OCT-1998; 98US-0103794.
XX
PR 30-APR-1999; 99US-0132068.
XX
PA (CHIR) CHIRON CORP.
XX
PI Frazier CM, Hickey E, Peterson J, Tettelin H, Venter JC;
PI Masignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;
PI Rappelli R, Pizzo M;
XX
DR WPI; 2000-318079/27.
XX
```

```
PT Isolated nucleotide sequences of Neisseria meningitidis which can be
PT used in the diagnosis and treatment of N. meningitidis infection and
PT other Neisserial infections, for example, N.gonorrhoea -
XX
PS Claim 7; Page 866-1272; 1760pp; English.
XX
CC The present invention describes methods of obtaining immunogenic
CC proteins from Neisseria genomic sequences. AAA81453 to AAA82414
CC represent specifically claimed Neisseria meningitidis genomic DNA
CC sequences; AAA81260 to AAA81303 and AAB25620 to AAB25663 represent
CC Neisseria DNA sequences and their corresponding proteins; AAA81254 to
CC AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the
CC isolation of Neisseria meningitidis DNA sequences; and AAA81322 to
CC AAA81452 represent Neisseria meningitidis MenB polynucleotide ORF
CC sequences, which are all used in the exemplification of the present
CC invention. The nucleic acid sequences, protein sequences, and antibodies
CC against them, can be used in the manufacture of a composition. The
CC composition can be used as a medicament (or in the manufacture of a
CC medicament) for treating, preventing or diagnosing infection due to
CC Neisserial bacteria. For example, some of the identified proteins could
CC be components of vaccines against Meningococcus B; against all serotypes;
CC and/or against all pathogenic Neisseriae. Identification of sequences
CC from the bacterium will also facilitate production of biological probes,
CC particularly organism-specific probes. Attempts to make efficacious
CC Meningococcus B vaccines have also been tried but none have successfully
CC overcome antigenic variability. The provision of further, complete
CC sequences may provide an opportunity to identify secreted or surface
CC exposed proteins that may be presumed targets for the immune system and
CC which are not antigenically variable or at least more conserved than
CC other more variable regions.
XX
SQ Sequence 1437668 BP; 344338 A; 353206 C; 385074 G; 355045 T; 5 other;
XX
Alignment Scores:
Pred. No.: 2.18e-06 Length: 1437668
Score: 222.00 Matches: 64
Percent Similarity: 43.44% Conservative: 32
Best Local Similarity: 28.96% Mismatches: 84
Query Match: 8.45% Indels: 42
DB: 21 Gaps: 6
US-10-088-045-2 (1-502) x AAA81490 (1-1437668)
QY 301 GlyGluAsnHisAsnIleGlnThrLeuIleProAlaGlyGlyAlaIleThrGlu 320
DB 748399 GAGCAACAACATCAAGTGCAG-----CCTTATATGAGTAAATTAAACACAG 748352
QY 321 PheLysValAspValProGlyAspTyr----- 329
DB 748351 TACGGGCTTGCGCTCGCTTG-CGGTACTATTGTACTGCTGCGGCTTGCCTTATC 748293
QY 330 ValIeuValAspHisAlaIlePheAsnIleGlyAlaLeuGlyIleLeuLys 349
DB 748292 CTGATTTTGTATTCACATCATCTACTGATTCGGCGGCAAAATCTTTCACCGCGTACGC 748233
QY 350 ValIGluGlyGluGlnAsnHisGluIleTyrSerHisIleGlnThr-----AspAlaVal 367
DB 748232 -----AAATTCACAACAACAGGTTTCAACATGAAAGAAACCAATGAACACACACC 748182
QY 368 TyrLeuProGlu-----GlyAlaProGlnAlaIleAspThr 379
DB 748181 CGACTGCCGACCGCGCTGCTTGGGCTGCTTGGCGCGCGCGCTTCTGCGCGGACAC 748122
QY 380 GlnGluAlaProLysThrProAlaProAlaAsnLeuGlnGluGlnIleLysAlaGlyLys 399
DB 748121 AGCATCATGACAA-----GGGCAA 748101
QY 400 AlaThrTyrAspSerAsnGlyAlaIleCysHisGlnProAspGlyLysGlyValProAsn 419
DB 748100 AAAGTGTACGAATCCACTGCTGCTGCGCGCAAAAGGCGAAAGCGCGGACACC 748041
QY 420 AlaPheProProLeuAlaAsnSerAspTyrLeuAsnAlaAspHisAlaArgAlaIleAsn 439
```

```
DB 748040 ATGTTTCCGCGCCCTACCGCTCCGACTTCATCATGAAAAAACCAGGTGCTGTGCAC 747981
QY 440 ILevalAlaenGlyLeuSerGlyLysIleThrValaenGlyAsnGlnTyrGluSerVal 459
DB 747980 AGCATGCGTCAAGGATCAACGGTACATCAAAAGTCAACGGCAAAACCTTACACGATTC 747921
QY 460 MetProAlaIleAlaLeuSerAspGlnGlnIleAlaAsnValIleThrTyrThrLeuAsn 479
DB 747920 ATGCCCGCAACCGCCATCAGCATCGGACATTCGCCGCGCTGCCACCTTATATCATGAC 747861
QY 480 SerPheGlyAsnLysGlyGlyGlnLeuSerAlaAspAspValAlaLysAlaLysLysThr 499
DB 747860 GCCTTTGACAAACGGCGCGGAGAACGCTTACGAAAAAGACGTAAAAACGACAAAAAGCMAA 747801
QY 500 Lys 500
DB 747800 AAA 747798

RESULT 17
AAZ53034
ID AAZ53034 standard; DNA; 341 BP.
XX
AC AAZ53034;
XX
DT 21-MAR-2000 (first entry)
XX
DE Neisseria meningitidis ORF 007 partial DNA sequence SEQ ID NO:39.
XX
KM Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
XX
OS Neisseria meningitidis.
XX
PN WO957280-A2.
XX
PD 11-NOV-1999.
XX
PF 30-APR-1999; 99WO-US09346.
XX
PR 01-MAY-1998; 98US-0083758.
PR 31-JUL-1998; 98US-0094869.
PR 02-SEP-1998; 98US-0098994.
PR 02-SEP-1998; 98US-0099062.
PR 09-OCT-1998; 98US-0103749.
PR 09-OCT-1998; 98US-0103796.
PR 09-OCT-1998; 98US-0103796.
PR 25-FEB-1999; 99US-0121528.
XX
PA (CHIR ) CHIRON CORP.
XX (GENO-) INST GENOMIC RES.
XX
PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
PI Petersen J, Pizzo M, Rappuoli R, Ratti G, Scalzo E, Scarfelli M;
PI Tettelin H, Venter JC;
XX
DR MPI, 2000-062150/05.
XX P-PDB; AAY74272.
XX
PT Novel Neisserial polypeptides predicted to be useful antigens for
PT vaccines and diagnostics
XX
PS Claim 7; Page 178-179; 1453pp; English.
XX
XX AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
XX represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
XX and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ54733 represent
XX PCR primers used in the exemplification of the present invention. The
XX polypeptides, the polynucleotides, antibodies and compositions of
XX the invention can be used as vaccines, as diagnostic reagents, and as
XX immunogenic compositions. The polypeptides can be used in the
XX manufacture of medicaments for treating or preventing infection due to
```

```
CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the
CC presence of Neisseria bacteria, or to raise antibodies. They may also
CC be used to screen for agonists or antagonists, which may themselves
CC have use as antibacterial agents. The polynucleotides of the invention
CC may also be used in gene therapy protocols.
XX
SQ Sequence 341 BP; 91 A; 113 C; 79 G; 58 T; 0 other;
Alignment Scores:
Pred. No.: 8.05e-08 Length: 341
Score: 184.00 Matches: 36
Percent Similarity: 55.95% Conservative: 11
Best Local Similarity: 42.86% Mismatches: 37
Query Match: 7.01% Indels: 0
DB: 21 Gaps: 0
US-10-088-045-2 (1-502) x AAZ53034 (1-341)
QY 398 GlyLysAlaThrTyrAspSerAsnGlyAlaAlaCySHisGlnProAspGlyLysGlyVal 417
DB 88 GGGCAAAAAGTGTACGATCCAACTGCGCTCGCTCCACGGCAAAAGGGCGAAGCGCGC 147
QY 418 ProAsnAlaPheProProLeuAlaAsnSerAspTyrLeuAsnAlaAspHisAlaArgAla 437
DB 148 GGAACCATGTTCCGCGCGCTTACCGCTCCGACTTCATCATGAAAAAACCAGGTGCTG 207
QY 438 AlaSerIleValAlaAsnGlyLeuSerGlyLysIleThrValaenGlyAsnGlnTyrGlu 457
DB 208 CTGCACAGCATGTGTCAAGGATCAACGGTACATCAATCAAAAGTCAACGGCAAAACCTACAC 267
QY 458 SerValMetProAlaIleAlaLeuSerAspGlnGlnIleAlaAsnValIleThrTyrThr 477
DB 268 GGATTCATGCGCGGACCGCATTCAGCATGCGCATTCGCCGCTGCCACTTATATATC 327
QY 478 LeuAsnSerPhe 481
DB 328 ATGAACGCTTT 339

RESULT 18
AAZ53035
ID AAZ53035 standard; DNA; 341 BP.
XX
AC AAZ53035;
XX
DT 21-MAR-2000 (first entry)
XX
DE Neisseria meningitidis ORF 007 partial DNA sequence SEQ ID NO:41.
XX
KM Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
XX
OS Neisseria meningitidis.
XX
PN WO957280-A2.
XX
PD 11-NOV-1999.
XX
PF 30-APR-1999; 99WO-US09346.
XX
PR 01-MAY-1998; 98US-0083758.
PR 31-JUL-1998; 98US-0094869.
PR 02-SEP-1998; 98US-0098994.
PR 02-SEP-1998; 98US-0099062.
PR 09-OCT-1998; 98US-0103749.
PR 09-OCT-1998; 98US-0103796.
PR 09-OCT-1998; 98US-0103796.
PR 25-FEB-1999; 99US-0121528.
XX
PA (CHIR ) CHIRON CORP.
XX (GENO-) INST GENOMIC RES.
XX
PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
```

PI Petersen J, Pizsa M, Rapuoli R, Ratti G, Scalato E, Scarselli M,  
PI Tettelin H, Venter JC;  
XX WPI; 2000-062150/05.  
DR P-PSDB; AAY74273.  
XX  
XX Novel Neisserial polypeptides predicted to be useful antigens for  
PT vaccines and diagnostics  
PS  
XX Claim 7; Page 179; 1453pp; English.  
XX  
XX AA253015 to AA254536, AA254577 to AA254615, and AAY74253 to AAY75941  
CC represent novel Neisseria meningitis and N. gonorrhoeae polynucleotides  
CC and polypeptides. AA254537 to AA254576 and AA254616 to AA25473 represent  
CC PCR primers used in the exemplification of the present invention. The  
CC polypeptides, the polynucleotides, antibodies and compositions of  
CC the invention can be used as vaccines, as diagnostic reagents, and as  
CC immunogenic compositions. The polypeptides can be used in the  
CC manufacture of medicaments for treating or preventing infection due to  
CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the  
CC presence of Neisseria bacteria, or to raise antibodies. They may also  
CC be used to screen for agonists or antagonists, which may themselves  
CC have use as antibacterial agents. The polynucleotides of the invention  
CC may also be used in gene therapy protocols.  
XX  
XX SQ Sequence 341 BP; 90 A; 114 C; 79 G; 58 T; 0 other;  
Alignment Scores:  
Pred. No.: 8,05e-08 Length: 341  
Score: 184.00 Matches: 36  
Percent Similarity: 55.95% Conservative: 11  
Best Local Similarity: 42.86% Mismatches: 37  
Query Match: 7.01% Indels: 0  
DB: 21 Gaps: 0  
US-10-088-045-2 (1-502) x AA253035 (1-341)  
QY 398 GILYSAATHrTYrAspSerAnCySAIAAaCYeHieGlinProAspGlyLyseGlyVal 417  
DB 68 GGGCAAAAAGTATGCAATCAACTGCTGCTGCGCAAGGCAAGGCGCC 147  
QY 418 ProAnaAlaPhProProLeuAlaAsnSerAspTYrLeuAsnAlaAspHisAlaArgAla 437  
DB 148 GGAACCATGTTCGCGCGCTTACCGCTCGACTCATCATGAAAAACCGCAGTGTCTG 207  
QY 438 AlaSerIleValAlaAsnGlyLeuSerGlyLyIleThValAsnGlyAsnGlyTyrGlu 457  
DB 208 CTGCACAGCATGTCAAAAGGATCAACGATCAATCAAGTCAACGCAAAACCTACAC 267  
QY 458 SerValMetProAlaIleAlaLeuSerAspGlnGlnIleAlaAsnValIleThrTYrThr 477  
DB 268 GGATTCATGCGCGCCACATCGCATCGGATCGGAGCATTCGCCCGCTCCACTTATATC 327  
QY 478 LeuAsnSerPhe 481  
DB 328 ATGAACGCTTT 339  
RESULT 19  
ABLS2917  
ID ABLS2917 standard; DNA; 4290 BP.  
XX  
XX ABLS2917;  
AC  
XX 16-JUL-2002 (first entry)  
DT  
XX 2-keto-D-gluconate dehydrogenase coding sequence.  
DE  
XX Cell membrane bound; 2-keto-D-gluconate dehydrogenase; enzyme;  
KW 2,5-diketo-D-gluconate; gene; ds.  
XX  
OS Unidentified.  
XX  
FH Key Location/Qualifiers

FT CDS 337..3991  
FT /tag= a  
FT /product= "2-keto-D-gluconate dehydrogenase"  
FT /note= "Contains one intron"  
FT exon 337..900  
FT /tag= b  
FT /number= 1  
FT intron 901..926  
FT /tag= c  
FT /cont\_splICE= (5'site:NO,3'site:NO)  
FT /number= 1  
FT exon 927..3988  
FT /tag= d  
FT /transl\_except= (pos:2580..2581, aa:Met)  
FT /number= 2  
PN KR2000019366-A.  
XX  
XX PD 06-APR-2000.  
XX  
XX PF 10-SEP-1998; 98KR-0037413.  
XX  
XX PR 10-SEP-1998; 98KR-0037413.  
XX  
XX (KOAD ) KOREA ADV INST SCI & TECHNOLOGY.  
XX  
XX Shin YC, Bahn JG, Yeom DY;  
PI  
XX  
XX WPI; 2001-088033/10.  
XX P-PSDB; AAM48440.  
XX  
XX Base sequence of novel cell membrane-bound 2-keto-D-gluconate  
PT dehydrogenase gene and method for conversion into  
PT 2,5-diketo-D-gluconate - Noabstract  
XX  
XX PS Claim 2; Page 6-8; 16pp; Korean.  
XX  
XX The present invention relates to a novel cell membrane-bound  
CC 2-keto-D-gluconate dehydrogenase and a method for conversion into  
CC 2,5-diketo-D-gluconate. The present sequence is the coding sequence for  
CC the 2-keto-D-gluconate dehydrogenase.  
XX  
XX SQ Sequence 4290 BP; 992 A; 1177 C; 1226 G; 895 T; 0 other;  
Alignment Scores:  
Pred. No.: 1.56e-05 Length: 4290  
Score: 175.00 Matches: 109  
Percent Similarity: 36.10% Conservative: 52  
Best Local Similarity: 24.44% Mismatches: 163  
Query Match: 6.66% Indels: 124  
DB: 23 Gaps: 20  
US-10-088-045-2 (1-502) x ABLS2917 (1-4290)  
QY 134 AspSerIleMetProHisAsnValAspPheHisAlaIleThrGlyProGlyGly----- 151  
DB 2620 GACGTTAAGCGCTCCGCGAGACAGACCGATGTGACATTTGCGCCGAGGGAACAGGT 2679  
QY 152 -----Gly-AlaGluAlaSerPheThrAlaProGlyHisIleSerThrPheSe 167  
DB 2680 CGCCATGCTTCGATGTGTCAGGATGCGACACCGCTCCGCGAGTGCACACCGCTTCG 2739  
QY 167 r-----PheValAlaLeuGlnProGlyLeuTyrValTyrHisCysAlaValAlaPro-- 184  
DB 2740 CGCGGTTATGATGATGCTCGCGATGATGATGATGATTCACCAATATCACCCTTC 2799  
QY 185 -----ValGlyMetHis-----IleAlaAsnGlyMet 193  
DB 2800 TCCCGACGCGATCGGCGCTTACAGGAGCATGATTTTCGACGGCGTACGCGATGCGCT 2859  
QY 193 tTyrGly-----LeuIleLeuValGlnProGlySerGlyLeuProGlyValAs 209  
DB 2860 ACGCGCTGACGCGCGCGATCTTATCCGCGATGCTTACACCTTACAGCAAGATCAC 2919







```

Db      906 CATCCCGCGCGTCCGGAAGCAAGAACTTCGT-----CAGATGACGGTCAAGAC 956
Qy      385 rProAlaProAlaLeuIngluInlIleValAlaGlyValAlaThrTyraPserAe 405
Db      957 GACCGCTCTGTGTGAACAAGGTGGCCAGGCGCAAGGTGGGAAGTCTATCTGCACAA 1016
Qy      405 nCyAlaAlaCyAlaIngluInlIleValAlaGlyValAlaThrTyraPserAe 425
Db      1017 CTGGCGCATCTGCGACATGAACGATGTGACGGGCTCAACCGCATGTTCCCGCGCTGCG 1076
Qy      425 a---AsnSerAspTyrlleuAsnAlaAspPhAlaArgAlaAlaSerIleValAlaAsnG 444
Db      1077 AGGCACCCGCGTCTCATTCACGACATCCGACCTGCGCTCGCATATGCTAGCCTTCGG 1136
Qy      444 YLeuSerGlyLysIleThrValaIngluInlIleValAlaGlyValAlaThrTyraPserAe 463
Db      1137 T-----GGCATCTGCGCGCGCGACCAACAGCGACCGCTCGCTGTCATGCGCGGTT 1190
Qy      463 eaLa-----LeuSerAspGlnGlnIleAlaAsnValIleThrTyrlleuAsnSerPh 481
Db      1191 CAGAATCACCCTCTCGACACGAGATGGCGATGTGTAACCTTTATGGCGAAGGCTG 1250
Qy      481 eGlyAsnIleGly---GlyGlnLeuSerAlaAspAspValAlaIleValAlaLysIleThr 499
Db      1251 GGGCAACAGCAACGCGGAACCGTGTCCGCTTCGATATCCAGAACTGCGCACAGCG 1308

```

## RESULT 22

AAQ48234 standard; DNA; 1437 BP.

AAQ48234;

13-JAN-1994 (first entry)

Cytochrome C gene.

C-553; improved produ.; production; oxidative fermentation; ss.

Gluconobacter suboxydans IFO 12528.

Key Location/Qualifiers

FT sig\_peptide 1..108

FT mat\_peptide 109..1437

FT /\*tag= a

FT /\*tag= b

PN JP05049480-A.

PD 02-MAR-1993.

PF 14-JUN-1991; 91JP-0238579.

PR 14-JUN-1990; 90JP-0154096.

PA (ASAHI ) ASAH I CHEM IND CO LTD.

DR WPI; 1993-169634/21.

XX P-PSDB; AAR39355.

XX

XX

XX

XX

XX

XX

```

Pred. No.: 4.35e-05 Length: 1437
Score: 163.00 Matches: 49
Percent Similarity: 49.26% Conservative: 18
Best Local Similarity: 36.03% Mismatches: 59
Query Match: 6.21% Indels: 10
DB: 14 Gaps: 6
US-10-088-045-2. (1-502) x AAQ48234 (1-1437)
Qy      369 LeuProGluGlyAlaProGlnAlaIleAspThrGlnGluAlaProIleThrProAlaPro 388
Db      916 GTGCCGGAAGGCAAGAACTTCGT-----CAGATGACGGTCAAGACCGCTTCG 966
Qy      389 AlaAsnLeuGlnGluInlIleValAlaGlyValAlaThrTyraPserAsnCyAlaAla 408
Db      967 CTGAACAAGGTCGCCAGGCAACGAGGTGCGGAAGTCTATCTGCACAACTGCGCATC 1026
Qy      409 CyAlaIngluProAspGlyLysGlyValProAsnAlaPheProProLeuAla---AsnSer 427
Db      1027 TGCCACATGACGATGTGACGGGCGTCAACCGCATGTTCCCGCGTGGCAGGCAACCG 1086
Qy      428 AspTyrlleuAsnAlaAspPhAlaArgAlaAlaSerIleValAlaAsnGlyLeuSerGly 447
Db      1087 GTCTGATTAACGACGATCCGACGTGCTGCGCAATGTGTAGCTTCGT-----GGC 1140
Qy      448 LysIleThrValaIngluInlIleValAlaGlyValAlaThrTyraPserAe 464
Db      1141 ATCTGCGCGCGCGACCAACGCGACCGCTGCTGTCATGCGCGGTTCAAGATCAC 1200
Qy      465 LeuSerAspGlnGlnIleAlaAsnValIleThrTyrlleuAsnSerPheGlyAsnIle 484
Db      1201 CTCTCGACACAGAGATGGCGATGTGTAACCTTTATGGCGAAGGCTGGGCAACAC 1260
Qy      485 Gly---GlyGlnLeuSerAlaAspAspValAlaIleValAlaLysIleThr 499
Db      1261 GCACCGGAACCGTGTCCGCTTCGATATCCAGAACTGCGCACAGCG 1308

```

## RESULT 23

ABA96893 standard; DNA; 1437 BP.

ABA96893;

07-MAY-2002 (first entry)

Gluconobacter oxydans sorbitol dehydrogenase subunit 2 DNA, SEQ ID:7.

Sorbitol dehydrogenase; SDH; subunit 2; Gluconobacter suboxydans;

KM coenzyme PQQ dependent; haem group; ds.

OS Gluconobacter oxydans.

PN KR98069057-A.

PD 26-OCT-1998.

PF 26-FEB-1997; 97KR-0005929.

PR 26-FEB-1997; 97KR-0005929.

PA (KOAD ) KOREA ADV INST SCI &amp; TECHNOLOGY.

DR WPI; 1999-608154/52.

XX Chol US, Lee SG, Lee EH;

XX

XX

XX

XX

XX

The invention relates to a sorbitol dehydrogenase (SDH) from Gluconobacter oxydans (referred to as Gluconobacter suboxydans in the specification) and the gene encoding it. The sorbitol dehydrogenase

Alignment Scores:

CC comprises 3 subunits of 75 kD, 50 kD and 29 kD, utilizes coenzyme PQQ and has a hem group. The present sequence represents DNA encoding the 50 kD subunit (subunit 2) of Gluconobacter oxydans sorbitol dehydrogenase.

XX Sequence 1437 BP; 296 A; 442 C; 418 G; 281 T; 0 other;

# Alignment Scores:

Pred. No.:	4,35e-05	Length:	1437
Score:	163.00	Matches:	118
Percent Similarity:	35.85%	Conservative:	34
Best Local Similarity:	27.83%	Mismatches:	144
Query Match:	6.21%	Indels:	130
DB:	20	Gaps:	25

US-10-088-045-2 (1-502) x ABA96893 (1-1437)

```

QY 171 LeuGlnProGlyLeuTyValTyHis-----CysAlaValAlaPro 184
DB 130 ATTCACGCGCGTCTACGTCGCGCCGCTGTCTGACTGCGTGCCTACCGCCTA 189
QY 185 ValGlyMetHisIleAlaAnglyMet-----TyrGly 195
DB 190 CACGCGCAGCCTTTTGTGTGTGTGTGAGATCAAGAGCCGATCGGCAGATCTACTCC 249
QY 196 Leu---IleuValGluProGlyGluGlyLeuProGlyValAspGlyTyrVal 214
DB 250 ACCAATCATCAGCGCTGACCCGAAATACGATATCGGCAC-----TATACA 294
QY 215 MetGlnGlyAspPheTyrThrTyGlyGlyTyGlyGlyGlnGlnGlnProPheAsp 234
DB 295 CTCGAAGAT-----TTCAGGAAGCG----- 315
QY 235 MetGlyLeuAlaIleAspGluAsp----- 242
DB 316 ATCGTAAAGGATACCGCAAGAGAGCGCGCAGCGTTATCCGCCATCGGATCTCTAG 375
QY 243 -----AlaGlyTyrValValPheAspGlySerVal 252
DB 376 TTCGTCGCGCTGTGTATGACGATCAAGCCATGATATCCCTTTCATGATGCGGTG 435
QY 253 GlyAla-----LeuThrGlyLeuAlaLeu 261
DB 436 AA-GCCGTCGCGCTTCAGAAAGCAGCGGCATCTCTGCGCGCATGAACATGCGCTG 494
QY 262 LysAlaLysValGlyLeuThrValArg-LeuPheValGlyAsnGlyGlyProAsnLeu 281
DB 495 -----GCCGTCGCGCATCTGCGCGCATGTTGTTCCGATGTCAACAGGCGCTCGA 548
QY 281 rSerSerPheHis-----ValIleGlyLeuIlePheAspLysValHisPheG 297
DB 549 CAAGACATCTCCGATCCGAGAGAGTGGCGCGGTGGGAAATACCTC-----GTAAATGCGCC 602
QY 297 uGlyGlyLysGlyGlyAsnHisAsnIleGlnThrLeuIleProAlaGlyAlaAl 317
DB 603 AGGCCATTTGGCGAGGTGATACG-----CCCGTCGCGATGCGCAT 644
QY 317 aIleThrGluPheLysValAspValProGlyAspTyrValLeu-----Va 332
DB 645 GCAGGTCAAGGGTATACGCCCAAGACGCAACGCTTACCTCCGCGTGGCGCACCGAT 704
QY 332 IaerHisAlaIle-----PheArgAlaPheAsnLysGlyAlaLeuGly----- 346
DB 705 CGAACAACTGATGTCCTCCAGCTGCGTGTAGCAATAGAGACAGCGGTGTGGTGTGTC 764
QY 347 -----IleuLysValGlyGlyGlyAsnHisGlyLeuTyrSer 360
DB 765 TGAAGACGATTCGCGAGTTCTGAAG--AGGGCGCGATGACATTCGCGCTT 821
QY 360 rHisLysGlnThrAspAlaVal---Tyr-LeuProGlyGlyAlaProGlnAlaIleAsp 379
DB 822 CGGTGCGATGTCGACGTGTGTGCTTACACACCCAGACCTGAGCCGACGATCTCGA 881

```

```

QY 379 hrcGlnGluAla-----ProLysThrProAlaProAlaAsnLeu--G 392
DB 882 CGCAAG-GCCAAAGTACTGAAGAGCATGCGCGCTTCGGAAGGCAAAAACCTGGTGC 940
QY 392 lngGlnIleLys-----AlaGlyLysA 400
DB 941 AGGATACGCGAGGACGCGCCCTGCTCGAAGCCGCTGGCAAGGTGATGCGAGCGCAG 1000
QY 400 lathrTyraSerPheAsnGlyAlaCysHisGlnProAspGlyLysGlyValProAsn 420
DB 1001 AGGTTTAACTTCCCAACATGTCCTGATGCAATGACGATGCTGTGTACACCGCA 1060
QY 420 IaPheProGlyLeuAlaAsnSerAspTyrLeuAlaAlaAspHisAlaArgAla--Ala 439
DB 1061 TGTTCGCGCGCTGCTGTCGCAACCGCGTGTCTATCAACGACATGCAACTCAAGGCCA 1120
QY 439 erIleValAlaAnglyLeuSerGlyLysIleThrValAsnGlyAsnGlnTyrGlySer 459
DB 1121 ACATCGTGAACATTCGCG-----GGTATCTGCTCCGACGAATACGCGCCATCTGCTG 1174
QY 459 aI---MetProAlaIleAla-----LeuSerAspGlnGlnIleAlaAsnValIleThr 476
DB 1175 TTGCGATGCGCGGCTTCCGCGATCATCTGTGTGACGAGATGCGCGATGTGTGACT 1234
QY 476 yThrLeuAsnSerPheGlyAsnLysGly---GlyGlnLeuSerAlaAspAspValAla 495
DB 1235 TCATGCGCAAGAGCTGGGCGCAACGAGCTCCGGAACCTGTGTGCTGGATATCCGA 1294
QY 495 yAlaLys 497
DB 1295 AGCTCCGC 1302

```

## RESULT 24

```

ID AAC83154 standard; DNA; 1437 BP.
AC AAC83154;
XX 01-MAR-2001 (first entry)
XX DE Sorbitol dehydrogenase subunit 2 DNA sequence.
XX KM Sorbitol dehydrogenase; SDH; Gluconobacter oxydans; D-sorbitol;
XX KW L-sorbitol production; 2-keto-L-gulononic acid; ds.
XX OS Gluconobacter oxydans.
XX PN W0200065066-A1.
XX PD 02-NOV-2000.
XX PF 23-APR-1999; 99WO-IB00736.
XX PR 23-APR-1999; 99WO-IB00736.
XX PA (CHOI/) CHOI E.
XX PA (RHEE/) RHEE S.
XX PA (LEE/) LEE E.
XX PI Choi E, Rhee S, Lee E;
XX DR WPI; 2000-687351/67.
XX DR P-PSDB; AAB35988.
XX PT Novel membrane-bound sorbitol dehydrogenase nucleic acid molecule
XX PT isolated from Gluconobacter suboxydans useful for the fermentative
XX PT production of 2-keto-L-gulononic acid and L-sorbitol from D-sorbitol
XX PS Claim 6; Fig 8; 96pp; English.
XX CC This invention relates to an isolated membrane-bound sorbitol
XX CC dehydrogenase (SDH) from Gluconobacter oxydans. The invention includes
XX CC polynucleotides (AAC83153 - AAC83154) encoding the 3 subunits of SDH

```



PS Claim 1; Fig 2; BPP; Japanese.

XX The sequence was obt'd. from a clone isolated from a library prepd.  
CC from RNA extracted from the fruit of *C. moschata*. The DNA can be  
CC used to produce the enzyme in high yields.

XX Sequence 2064 BP, 574 A, 444 C, 493 G, 553 T, 0 other;

#### Alignment Scores:

Pred. No.:	Length:	Matches:	Conservative:
Score:	7,07e-05	2064	83
Percent Similarity:	163.00	163.00	44
Best Local Similarity:	37.69%	24.63%	130
Query Match:	6.21%	80	19
DB:	12	Gaps:	19

US-10-088-045-2 (1-502) x AA014178 (1-2064)

```

QY 101 ValGluTyrGlnPheThr-----PheGlyGly 110
DB 118 GTCGAGTACATGTTGGGCTCCTGATGTGTAATAATATCGTATGGGAATCAAGCGC 177
QY 111 GlnValProGlyGlnMetIleArgValArgGlyAspThrIleGluValGlnPheSer 130
DB 178 CAGTTTCCCTGAGCACTACGATTCAGGCCAACGCTGTGATACTGTTGTTGTCGCTGAT 237
QY 131 Aen-----HisProAspSerLysMetProHisAenValAspPheHisAlaIleThrGly 148
DB 238 AATTAAGCTCCACACTGAAAGTGTGTCATTCAC-----TGGCATGGAATTTTGCA 288
QY 149 ProGly-----GlyGlyAlaGluAlaSerPheThrAla-----ProGly 161
DB 289 CGAGGAATCCGTCGCTGATGTCACATGCTTCATCTCCAGTGTCTATTAACTGCTGGT 348
QY 162 HisThrSerThrPheSerPheLysAlaLeuGlnProGlyLeuTyrValTyrHisCysAla 181
DB 349 GAACGCTTCTTACAACTTTCATCTGTGTAACCCCTGGAGCGTTTCTATCAT----- 402
QY 182 ValAlaProValGlyMetHisIleAlaAngLysMetTyrGlyLeuIleLeuValGlnPro 201
DB 403 ---GGTCATTTGGGAATGCCAAGATGCCAGGGGTATATGATCTTTGATGAGATCCT 459
QY 202 LysGlnGlyLeuProLys-----ValAspLysGlnTyrTyrValMetGlnGlyAsp 218
DB 460 CCACAAGAGAAAAGAGCCATTCATTTAGACGAGAAATACAACTATTACCACTGAT 519
QY 219 PheTyrThrLysGlyLysTyrGlyGln---GlyLeuGlnProPheAspMetGlyLys 237
DB 520 TGGTGCATCAAAAGTATTCATTAACAAAGAGTTGCTTACG-----TCCAAA 567
QY 238 AlaIleArg-----GluAspAlaGluTyrValValPheAsnGly----- 250
DB 568 CCTATTCTGTCGATCGGTGAGGCTCAACCATATCTTTAAATGGAAGAGGCGAGTTGCAT 627
QY 251 ---SerValGlyAlaLeuThrGlyGluAsn-----AlaLeuLysAlaLysValGlyGlu 267
DB 628 TGTTCATTTGACGCTAAATACAGACAGTATTAGACCCTATTAATAAGAAAGTGA 687
QY 268 ThrValArgLeuPheValGlyAsnGlyGlyProAsnLeuThrSerPheHisValIle 287
DB 688 CCGTGGCAGCTTATATC-----TTCCATGTTAAG 717
QY 288 GlyGluIlePheAspLysValHisPheGluGlyLysGlyGluAsnHisAsnIleGln 307
DB 718 CCCAAAAGACTTACGAGATTAAGAAAT-----GCAAGT 750
QY 308 ThrThrLeuIleProAlaGlyAlaIleAlaIleThrGlu-----PheLysValAsp 324
DB 751 ACCACTGCTTTAGCGGCCCTCAACTTGCATTTGGGAATCAACCATTTGTTGGTGAAGA 810
QY 325 ValProGlyAspTyrVal-----LeuValAspHisAlaIlePheArgAlaPhe 340
DB 811 GCCAGCGTAATTAACGTCACCAACCATTTTACACCTCCGACATCGACATTTATTCGGCGAG 870

```

```

QY 341 AsnLysGlyAlaLeuGlyIleLeuLysValGluGlyGluAsnHisGluIleTyr--- 359
DB 871 TCTTACTCGCTCTCATACACACGACCAAAACCATCGGAGAACTACTGGGTCTCCGTC 930
QY 360 -----SerHisLysGlnThrAsp-----AlaValTyrLeuPro 370
DB 931 GGCACCCCGCGAGGAGCTCCCAACACACCGCGGATTTGACCTCTCAACTAATCACTAAC 990
QY 371 GluGlyAlaProGlnAlaIleAspThrGlnGluAlaProLysThrProAla 387
DB 991 AATTCGCTCGAAA---TTACCACTTCTCCGCGCGGAAACCCCGCC 1038
RESULT 26
ABA96894
ID ABA96894 strand; DNA; 4830 BP.
AC ABA96894;
XX
XX 07-MAY-2002 (first entry)
XX
DE Gluconobacter oxydans sorbitol dehydrogenase subunit 1-3-encoding DNA.
XX
XX Sorbitol dehydrogenase; SDH; subunit 1; subunit 2; subunit 3;
KW Gluconobacter suboxydans; coenzyme PQQ dependent; haem group; ds.
XX
XX Gluconobacter oxydans.
XX
XX KR98069057-A.
XX
XX 26-OCT-1998.
XX
XX 26-FEB-1997; 97KR-0005929.
XX
XX 26-FEB-1997; 97KR-0005929.
XX
XX (KOAD ) KOREA ADV INST SCI & TECHNOLOGY.
XX
XX ChOI US, Lee SG, Lee EH;
XX
XX WPI, 1999-608154/52.
XX
XX Sorbitol dehydrogenase of Gluconobacter suboxydans and gene thereof -
XX
XX Example; Page 14-18; 18pp; Korean.
XX
XX The invention relates to a sorbitol dehydrogenase (SDH) from
XX Gluconobacter oxydans (referred to as Gluconobacter suboxydans in the
XX specification) and the gene encoding it. The sorbitol dehydrogenase
XX comprises 3 subunits of 75 kD, 50 kD and 29 kD, utilizes coenzyme PQQ and
XX has a haem group. The present sequence represents DNA encoding all
XX 3 subunits of the Gluconobacter oxydans sorbitol dehydrogenase of
XX the invention.
XX
XX SQ Sequence 4830 BP, 1071 A, 1424 C, 1350 G, 985 T, 0 other;
XX
XX Alignment Scores:
XX
XX Pred. No.: 0.000221 Length: 4830
XX
XX Score: 163.00 Matches: 146
XX
XX Percent Similarity: 34.40% Conservative: 48
XX
XX Best Local Similarity: 25.89% Mismatches: 187
XX
XX Query Match: 6.21% Indels: 186
XX
XX DB: 20 Gaps: 33
XX
XX US-10-088-045-2 (1-502) x ABA96894 (1-4830)
QY 43 LysThrAlaAsn-----AlaAspAsnAlaIleAspGlnGluHisGlnGlyGluPro 59
DB 2804 AAGCGCGCAAGAGATCTCTACGACAGCAAGTCAGAGCCGGAAGAATAGAGCAGCGCTC 2863
QY 60 ProValIleAspAlaIleValThrHisAlaProGluValPro-----ProPro 75
DB 2864 CTAACGGGCAATTCCTC-----AACGTCCCTCAGGGTTCCGTTGCGCT 2908

```

Qy	76	ValAspArgSerHisProAlaIysValValValIleMetCysIuThrValGluIysValMet	95
Db	2909	GCACGCCGGGACCCATTCGGTAAGGGAACCGTC-----ACGCTGAAGAGATGACG	2955
Qy	96	ArgIleuAlaSerGlyValGluIuTrpHisPheGluYglIysIleAla-----	112
Db	2960	TGACATGCTCAGGCAATATC-----TCGGACACAGCTGGATCTGA	3007
Qy	113	-----ProGlyIleMetIleArgValArgGluGlyAspThrIleGluValGlnPheSer	130
Db	3002	GATGAACACAGGATGGAAATATCCGGCCGC-----	3033
Qy	131	AsnHisProAspSerIuMetProHisAsnValAspPheHisAlaAlaIleThrIuProGly	150
Db	3032	-----ACTCGGCTCATGGCAGCTGTCTTCGGTGTCCGCAAGCCAGGA	3076
Qy	151	GlyGlyAlaGluAlaSerPheThrAlaProGlyHisIleThrSerThrPheSerPheIysAla	170
Db	3077	C--GCTGATGACGCC-----CMG	3092
Qy	171	LeuGlnProGlyLeuIuTrpValTyrlHis-----CysAlaValAlaPro	184
Db	3093	ATTCACGCGGCTCCCTACAGCTGGCCGCCCTGTGACTGCTGCTCCGTCATACCCGACTA	3152
Qy	185	ValGlyMetHisIleAlaAsnGlyMet-----TyrlGly	195
Db	3153	CACGGCCAGCCTTTCTGCTGCTGTGGATGATCAAGCCGATGGGACGACATCACTACC	3212
Qy	196	Leu---IleLeuValGlnProIysGluGlyLeuProIysValAspIysGluIuTrpVal	214
Db	3213	ACCAACATACCGCTGACCGGAAATACGGATATCCGCACAC-----TATACA	3257
Qy	215	MetGlnGlyAspPheIuTrpThrIuysGlyIuStyrlGlyGluGlnGlyLeuGlnProPheAsp	234
Db	3258	CTCGAAGAT-----TTCAAGAGCG-----	3278
Qy	235	MetGluIysAlaIleArgGluAsp-----	242
Db	3279	ATCCGTAAGGATATCCGCAAGACGCGACGCTTATCCGGCCATCCGTATCTTGAG	3338
Qy	243	-----AlaGluIuValAlaPheAsnGlySerVal	252
Db	3339	TTCCGCTCGGCTCTGTGATGACACATCAAGGCCATGTATGCTTCTTATGCAATGGCGTG	3398
Qy	253	GlyAla-----LeuThrGlyIuAsnAlaIeu	261
Db	3399	AA-GCCGGTGCCTTCAGAACACAGCCGACATCTCTGGCCGAGAACACGCGCTG	3457
Qy	262	IysAlaIysValGlyIuThrValArg-LeuPheValGlyIuysnGlyIuProAsnIeuThr	281
Db	3458	-----GCCGTTGGCCATCTGGCGCGCATGTGTGTTCCAGTGTCAACACGAGGCCCTGA	3511
Qy	281	rSerSerPheHis-----ValIleGlyIuIlePheAspIysValHisPheG	297
Db	3512	CAGAGCATCTCCGATCCGGAAGTGGCGCGGAGATACCTC-----GTGATGGCCC	3565
Qy	297	uGlyGlyIysGlyIuysnHisAsnIleGlnIleThrIleuIleProAlaGlyIuAla	317
Db	3566	AGGCATTTGGCGGAGTGTCAACG-----CCCGGTGCATGGCCAT	3607
Qy	317	AlleThrGluPheIysValAspValProGlyAspIuValIeu-----Val	332
Db	3608	GCAGGTCAAGGCGTATACGGCCAGAGGACGCAAGCTTACTCTCCGGTGGCGACCGCAT	3667
Qy	332	IAspHisAlaIle-----PheArgAlaPheAsnIysGlyAlaIeuGly-----	346
Db	3668	CGACAACATGATGTCTCCACGCTGCGGTAGCAATAGACACACGGGTCTGGGTGCTGTGTC	3727
Qy	347	-----IleLeuIysValGluGlyIuGluIuysnHisGlyIuIeTrp	360
Db	3728	TGAAGACACATTCGCCAGTCTCTGAAG--ACGGCGCGTATGACCAATTCCTCCGCTTT	3784

Qy	360	hHsLysysGlnThrAspAlaVal--Tyr-LeuProGluGlyAlaProGlnAlaIleAspT	379
Db	3785	CGGTGCATGCTGTGAAGTGTGGCCCTTAAGCAACCCAGCACTGGACCGAGCATCTGCA	3844
Qy	379	hGlnGlnIleA-----ProLysThrProAlaProAlaAsnLeu---G	392
Db	3845	CCCAACG-GCCAAATACCTGAAGAGCATGCGCGCTTCCGAAAGCAAAAACCTGGGTG	3907
Qy	392	InGluGlnIleLys-----AlaGlyLysA	400
Db	3904	AGGAATGACGGCAAGGCCACAGCGCCCTGCTCGAAGCCGATGGCAAGGATGATGACGGCGAG	3963
Qy	400	laThrTyrAspSerAsnCysAlaIaIaCysHISgInProAspGlyLysGlyValProAsnA	420
Db	3964	AGCTTTAATCCTCCACAACACTGTGTCATCTGCATATGAAGATGGCACTGTGTCAACCCCA	4022
Qy	420	laPheProProLeuAlaAsnSerAspTyrLeuAsnAlaAspHisAlaArgAla--Alas	439
Db	4024	TGTTTCCCGCCGCTGGCTGGCAACCCGGATCGATCATCGAAGAAATGCAACCTCACTCAATGGCCA	4088
Qy	439	erIleValAlaAsnGlyLeuSerGlyLysIleThrValAsnGlyAsnGlnIrryGusery	459
Db	4084	ACATGTGTACATCTGGC-----GSAITTTCTGCTCGACGAATACGGCCCATCTGCTG	4137
Qy	459	al---MetProAlaIleAla-----LeuSerAspGlnGlnIleAlaAsnValIleThrT	476
Db	4138	TTGCCATCCCGGGCTTCCGCGATCATCTGTCTGAACAGACAGATCGCGATGTTGAACT	4197
Qy	476	yrThrLeuAsnSerPheGlyAsnLysGly---GlyGlnLeuSerAlaAspAspValAlaL	495
Db	4198	TCATGCGCAAGAGCTGGGGCAACAGAGCTCCGGGAACCTGTCTGCTCGATATCCGCA	4257
Qy	495	ysAlaLys	497
Db	4258	AGCTCCGC	4265
RESULT 27			
AAc83156	AAC83156 standard; DNA; 4830 BP.		
ID	AAC83156		
XX	AAC83156;		
AC			
XX			
DT	01-MAR-2001 (first entry)		
DE	DNA encoding Gluconobacter sorbitol dehydrogenase.		
XX			
KW	Sorbitol dehydrogenase; SDH; Gluconobacter oxydans; D-sorbitol;		
KW	L-sorbose production; 2-keto-L-gulononic acid; ds.		
XX			
OS	Gluconobacter oxydans.		
XX			
EN	WO200065066-A1.		
XX			
XX	02-NOV-2000.		
PF	23-APR-1999; 99WO-IB00736.		
XX			
PR	23-APR-1999; 99WO-IB00736.		
XX			
PA	(CHOI/) CHOI E.		
PA	(RHEE/) RHEE S.		
PA	(LEE/) LEE E.		
XX			
PI	Choi E, Rhee S, Lee E;		
XX			
DR	WPI; 2000-687351/67.		
DR	P-PsDB; AAB35987, AAB35988.		
XX			
PT	Novel membrane-bound sorbitol dehydrogenase nucleic acid molecule		
XX	isolated from Gluconobacter'guboxydans useful for the fermentative		
XX	production of 2-keto-L-gulononic acid and L-sorbose from D-sorbitol		
XX	Claim 16; Fig 8; 96pp; English.		

XX This invention relates to an isolated membrane-bound sorbitol  
CC dehydrogenase (SDH) from Gluconobacter oxydans. The invention includes  
CC polynucleotides (AAC83153 - AAC83156) encoding the 3 subunits of SDH.  
CC (AA835987 - AA835989). Also included in the invention are two  
CC polynucleotide sequences AAC83156 and AAC83157 which contain the subunit  
CC coding regions. The sorbitol dehydrogenase polynucleotide sequences are  
CC useful for producing L-sorbitol from D-sorbitol and for increasing the  
CC production of 2-keto-L-gulononic acid by transforming a host cell.  
CC especially Gluconobacter with the DNA and selecting the transformed host  
CC cell. Note: An additional 66 DNA sequences (fragments of the SDH subunit  
CC genes) encoding fragments of SDH are specifically claimed, however these  
CC sequences are not given separately in the specification but are included  
CC in sequences AAC83156 and AAC83157.

XX  
SQ Sequence 4830 BP; 1071 A; 1424 C; 1350 G; 985 T; 0 other;

## Alignment Scores:

Pred. No.:	0.000221	Length:	4830
Score:	163.00	Matches:	146
Percent Similarity:	34.40%	Conservative:	48
Best Local Similarity:	25.89%	Mismatches:	187
Query Match:	6.21%	Indels:	186
DB:	21	Gaps:	33

US-10-088-045-2 (1-502) x AAC83156 (1-4830)

QY 43 LysThrAlaAsn-----AlaSerAsnAlaLaserGlnGlnGlnGlnGlnLeu 59  
DB 2804 AACGGCGCCAAACGATCTTACGACGAGAAAGTCAAGGCCGGAAGATGAGGACGCGCTC 2863  
QY 60 ProValIleAlaIleValThrHisAlaProGluValPro-----ProPro 75  
DB 2864 CCTAACGGCGCAATCTCTC-----AACGTCCTCAGGTTGGTGGCGCTC 2908  
QY 76 ValAspArgAspHisProAlaLysValValLysMetGluThrValGluLysValMet 95  
DB 2909 GCMAACCCCGGACCATCCGTAAACGGGAAACCGTC-----ACGCTGAAGAGATGACG 2959  
QY 96 ArgLeuAlaAspGlyValGluValGlnPheThrPheGlyGlnVal-----112  
DB 2960 TGACATGCTCAAGGCAATTAC-----TCGGACACGACTGGTATCTGA 3001  
QY 113 -----ProGlyGlnMetIleArgValArgGluGlyAspThrIleGluValGlnPheSer 130  
DB 3002 GATGAACACGAGGATGGAATAACCGCGCCG-----3031  
QY 131 AsnHisProAspSerLysMetProHisAsnValAspPheHisAlaIleThrGlyProGly 150  
DB 3032 -----AGTCGGCTCATGCGACAGTGTCTTTCGGTGGTCCCAACCCGACGA 3076  
QY 151 GlyGlyAlaGluLaserPheThrAlaProGlyHisThrSerThrPheSerPheLysAla 170  
DB 3077 C-GCTGATGACGCC-----CTG 3092  
QY 171 LeuGlnProGlyLeuValTyrrHis-----CysAlaValAlaPro 184  
DB 3093 ATTCAAGCGCGGTCTTAAGTGGCCCGCTGTGACTGTGCTGGCTGCATACCGACGTA 3152  
QY 185 ValGlyMetHisIleAlaAsnGlyMet-----TyrrGly 195  
DB 3153 CACGGCCAGGCTTTTGTGTGTGTGCTGAGATCAAGAGCCGATCGGACAGATCTATCC 3212  
QY 196 Leu---IleLeuValGluProLysGluGlyLeuProLysValAspLysGluTyrrVal 214  
DB 3213 ACCAACATCAAGCTGCGGAAATACGATATCGGCAAC-----TATACA 3257  
QY 215 MetGlnGlyAspPheTyrrHisGlyLysTyrrGlyGlnGlnGlnLeuGlnProPheAsp 234  
DB 3258 CTGGAAGAT-----TTACAGAAAGCG-----3278  
QY 235 MetGluLysAlaIleArgGluAsp-----242

DB 3279 ATCCGTAAGGATATCCGCAAGACGCGGCGTTTATCCGCGCATCTCTGAG 3338  
QY 243 -----AlaGluTyrrValValPheAsnGlySerVal 252  
DB 3339 TTGCGTCCGCTGTGTGATACACATCAAGGCCATATGCTTCTTACATGACGCGCTG 3398  
QY 253 GlyAla-----LeuThrGlyGlnAlaLeu 261  
DB 3399 AA-GCCGGTGGCCTTCAGAACAGACGCGGACATCTCTGCGCATGAACATGCGCTG 3457  
QY 262 LysAlaLysValGlyGluThrValArg-LeuPheValGlyAsnGlyLysProAsnLeuThr 281  
DB 3458 -----GCCGTGGCCATCTGGCGCGATGTTTGTTCGACTGTCAACACGCGCTCGA 3511  
QY 281 rSerSerPheHis-----ValIleGlyGluIlePheAspLysValHisPheG 297  
DB 3512 CAAGACATCTCCGATCCGGAAGTGGCGGCTGGCGAATACCTC-----GTGATAGCCC 3565  
QY 297 uGlyGlyLysGlyGluAsnHisAsnIleGlnThrLeuIleProAlaGlyValAla 317  
DB 3566 AGGCCATGTGGGAGATGTCATACG-----CCCGTGGCATGGCGCAT 3607  
QY 317 alleThrGluPheLysValAspValProGlyAspTyrrValLeu-----Va 332  
DB 3608 GCAGTCAAGGCTATATACGCGCAAGACGCGCAACGCTTCTCCGATGGCGCACCGAT 3667  
QY 332 LAspHisAlaIle-----PheArgAlaPheAsnLysGlyAlaLeuGly-----346  
DB 3668 CGACAACTGATGATCTCCACGCTGCGTACATTAACATACACACGCGTGGTGGCTGCTG 3727  
QY 347 -----IleLeuLysValGluGlyGluGluAsnHisGluLysLeu 360  
DB 3728 TGAAGACGACATTCGCCAGATCTCTGAAG-----AGCGCCGATGACATTCGCGCTT 3784  
QY 360 rHisLysGlnThrAspAlaVal---Tyr-LeuProGluGluAlaProGlnAlaIleAsp 379  
DB 3785 CGGTGCATAGCTGACGTGTGGTCCCTACAGACCCGACCTGACCGACGACGATCTGCA 3844  
QY 379 hGlnGluAla-----ProLysThrProAlaProAlaAsnLeu---G 392  
DB 3845 CGCAACG-GCCAAATGATCTGAAGACGATGCGCGCTTCGGAAGGCAAAACCTGGCTC 3903  
QY 392 hGlnGlnIleLys-----AlaGlyLysA 400  
DB 3904 AGGATACGCGCAAGCGCATGCGCTGCGAAGCGGTGGCAAGGTGATGACGGCGAG 3963  
QY 400 lAthrTyrrAspSerAsnCyAlaAlaCysHisGlnProAspGlyLysGlyValProAsn 420  
DB 3964 AGGTTAATCTCCAACTGTGCTGCATGTGCATATGAACATGACCTGGTCAACCGCA 4023  
QY 420 lAheProProLeuAlaAsnSerAspTyrrLeuAsnAlaAspHisAlaArgAla---Ala 439  
DB 4024 TGTTCGCGCGCTGCTGGCAACCGCGTGTATCATCGACCAATGCAACTCAATGGCCA 4083  
QY 439 exIleValAlaAsnGlyLeuSerGlyLysIleThrValAsnGlyAsnGlnTyrrGlnSer 459  
DB 4084 ACHTCTGACATTCGCG-----GATATCTGCTCCGACGAATAGCGCGCATCTGCTG 4137  
QY 459 aI---MetProAlaIleAla-----LeuSerAspGlnGlnIleAlaAsnValIleThr 476  
DB 4138 TTGCGCATGCGGGCTTCGCGCATCTGTCTGACAGAGATGCGCGGATGTGTGAAC 4197  
QY 476 yrrThrLeuAsnSerPheGlyAsnLysGly---GlyLeuSerAlaAspAspValAla 495  
DB 4198 TCATGCGCAAGGCTGGGCAACAGGCTCCGGGAACCTGTGCTCGATATCCGCA 4257  
QY 495 yEAlaLys 497  
DB 4258 AGCTCCGC 4265

RESULT 28  
AAAF26419  
ID AAF26419 standard; DNA; 2007 BP.

```

XX AF26419;
AC
XX 02:MAY-2001 (first entry)
DT
XX Pseudomonas sp heavy metal transporter encoding DNA ORF04246.
DE
XX Heavy metal transporter; iron transporter; transgenic plant;
KM homeostasis regulator; heavy metal ion; trace element; soil pollution;
KW plant growth promoter; plant development; ds.
XX Pseudomonas sp.
OS
XX DE19934720-A1.
PN
XX 25-JAN-2001.
PD
XX 23-JUL-1999; 99DE-1034720.
PF
XX 23-JUL-1999; 99DE-1034720.
PR
XX (TIGR-) TIGR INST GENOMIC RES.
PA (QIUA-) QIAGEN GMBH.
PA (GBR-) GES BIOTECHNOLOGISCHE FORSCHUNG MBH.
PA (DKF2-) DKFZ DEUT KREBSFORSCHUNGSZENTRUM.
PA (MED1-) MEDIZINISCHE HOCHSCHULE HANNOVER.
XX MPI, 2001-160596/17.
DR
XX New DNA encoding iron or heavy metal transporters, useful for promoting
PT growth of plants on polluted soil -
PI Claim 3a; Page 14-15; 54pp; German.
PS
XX This invention describes novel DNA sequences (I) that encode iron or
CC heavy metal transporters (II). The invention also describes (1) a
CC recombinant expression vector (III) containing (I); (2) prokaryotic or
CC eukaryotic cells (III) transformed with (I) or (III); (3) production of
CC (II) by culturing (III); (4) (partial) expression products (IV) of (I)
CC and synthetic proteins or peptides with the same sequences; (5)
CC antibodies (Ab) specific for (IV); (6) hybridoma cells that produce
CC monoclonal Ab; and (7) transgenic plants that contain (III). The iron or
CC heavy metal transporters encoded by (I) transport heavy metal ions across
CC cell walls and regulate homeostasis of trace elements. (I), and their
CC fragments are useful for: (1) expression of (II); (2) as probes and
CC primers for detection, isolation and amplification of full length cDNA
CC sequences; and (3) producing transgenic plants. (III) are used to promote
CC growth, development and yield of plants, particularly leguminosae,
CC especially when growing in soil polluted by heavy metals injurious to
CC plants. They also improve homeostasis of iron and trace elements. Host
CC cells that express the iron or heavy metal transporters take up heavy
CC metals, so reduce pollution of soil and release iron and other trace
CC metals which improves soil quality and protect plants against pollutants
CC from the soil.
XX
XX Sequence 2007 BP, 472 A; 569 C; 594 G; 372 T; 0 other;
SQ

```



QY 444 GlyLeuSerGly 447  
DB 1150 AACATGGCCGCT 1161  
RESULT 29  
AA199683/C  
ID AA199683 standard; DNA; 4403765 BP.  
XX  
AC AA199683;  
XX  
DT 15-JUN-2002 (first entry)  
XX  
DE Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 2.  
XX  
KM Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;  
XX variation; epidemiology; patient treatment; epidemic monitoring; ds.  
OS Mycobacterium tuberculosis.  
XX  
PN US6294328-B1.  
XX  
PD 25-SEP-2001.  
XX  
PF 24-JUN-1998; 98US-0103840.  
XX  
PR 24-JUN-1998; 98US-0103840.  
XX  
PA (GENO-) INST GENOMIC RES.  
XX  
PI Fleischmann RD, White OR, Fraser CM, Venter JC;  
XX WPI; 2001-647261/74.  
XX  
PT Evaluating strain variation of Mycobacterium tuberculosis, comprises  
PT determining the nucleotide sequence of the strain at positions in the  
PT genome corresponding to positions where M. tuberculosis strains CDC  
PT 1551 and H37Rv differ.  
XX  
PS Claim 4: SEQ ID NO 2; 3pp + Sequence Listing; English.  
XX  
CC The invention relates to evaluating strain variation within and between  
CC different populations of the tuberculosis bacterial pathogen,  
CC Mycobacterium tuberculosis or related Mycobacterium by determining the  
CC nucleotide sequence of the first strain at positions in the complete  
CC sequence of the genome that correspond to positions that differ in the  
CC nucleotide sequences of M. tuberculosis strains CDC 1551 (AA199683) and  
CC H37Rv (AA199682). The method is useful for evaluating strain variation of  
CC M. tuberculosis and has valuable application in the fields of  
CC tuberculosis genetics, epidemiology, patient treatment and epidemic  
CC monitoring.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from USPTO  
CC at seqdata.uspto.gov/sequence.html?DocID=6294328B1.  
XX  
SQ Séquence 4403765 BP; 757105 A; 1447799 C; 1441301 G; 757371 T; 189 other;  
XX  
Alignment Scores:  
Pred. No.: 5.26 Length: 4403765  
Score: 158.50 Matches: 104  
Percent Similarity: 31.00% Conservative: 51  
Best Local Similarity: 20.80% Mismatches: 180  
Query Match: 6.04% Indels: 165  
DB: 22 Gaps: 23  
US-10-088-045-2 (1-502) x AA199683 (1-4403765)  
QY 16 SerAlaLeuMetLeuSerGlyCysSerAsnGlnAlaAspLysAlaIleGlnProLysSer 35  
DB 944001 AGCGGCTTTCGCTTACCGCTTGC-----GCTTCGAAGCCACGCGCA 943960  
QY 36 SerThValAspAlaAlaAlaLysThrAlaAsnAlaAspAsnAla--AlaSerGlnGlu 54  
DB 36 SerThValAspAlaAlaAlaLysThrAlaAsnAlaAspAsnAla--AlaSerGlnGlu 54

DB 943959 TCC-----GGCGCCCGCGGATGACCGCTGCATCGACCGCGGCGCGCGCGCG 943906  
QY 55 HisGlnGlyGluLeuProValIleAspAlaIleValThrHisAlaProGluValProPro 74  
DB 943905 CACAGTGGCGCA-----ACGTTACCGCCACGCTGAC-----CCCGAG 943867  
QY 75 ProValAspArgAspHisProAlaLysValValLysMetGluThrValGluLysVal 94  
DB 943866 CCGCGAGATCGACCTGGGTGGCGGATCGTACGACGCTG----- 943825  
QY 95 MetArgLeuAlaAspGlyValGluTyrGlnPheThrPheGlyGlnValProGly 114  
DB 943824 -----ACCTACCGCAACACCATCCCGCA 943801  
QY 115 GlnMetIleArgValArgGluGluValAspThrIleGluValGlnPheSerAsnHisProAsp 134  
DB 943800 CCACTATTCGGGCGCACCGTGGGGATGAGATTGTCTCGGTGACCAACCGTGGGT 943741  
QY 135 SerLysMetProHisAsnValAspPheHis-----AlaAlaThrGlyProGlyGly 152  
DB 943740 GAT-----CCGACGTGGGTGATTCGACACGCGCGCGTGGCAACGATGATGATGCG 943687  
QY 153 AlaGluAlaSerPheThr-----AlaProGlyHisThrSerThrPheSerPheLysAla 170  
DB 943686 ACCGAGCCCGCGCATCGAATCGGCGCGCGGTGACTTTCGATCCGCTCTCCGTG 943627  
QY 171 LeuGlnProGlyLeuTyrValTyrHisCysAlaValAlaProValGlyMetHisIleAla 190  
DB 943626 CCGGATCCGGGACCTACTGCGGCGCCATCCGACGTCGCTTCAAGCGCACAC----- 943573  
QY 191 AsnGlyMetTyrGlyLeuLeuValGluProLysGluGlyLeuProLysValAspLys 210  
DB 943572 ---GGCTATATCTGCTGCTGCGTCGTCGACGATCCGACGAGCGACGCGCATCGACGCC 943516  
QY 211 GluTyrTyrValMetGlnGlyAspPheTyrThrLysGly----- 223  
DB 943515 GATGATGATCATCATCTCTTCACGAT---TGACCGAGCGGATCGGAGATCCCGCAACAG 943459  
QY 224 LysTyrGlyGluGlnGlyLeuGlnProPheAspMetGluLysAlaIleArgGluAspAla 243  
DB 943458 CTCTACGGGAGCTG-----ACCGACCGCAACCAACCATGCAACACACA 943411  
QY 243 ----- 243  
DB 943410 ACAGTATCCCGAAGCGAAGCGCTTGACAGCAACCTGCGCGCGCGACGAGGGAGAC 943351  
QY 244 -----GluTyrValValPheAsnGlySerValGlyAlaLeuThrGlyGluAsnAla 260  
DB 943350 ATCGCCTACCCGACTATCTGATCAACGGCGCATCC-----GTGGCGCGCACGCTC 943297  
QY 261 LeuLysAlaLysValGlyGluThrValArgLeuPheValGlyAsnGlyGlyProAsnLeu 280  
DB 943296 TTTAAGCGCAAGCTGGCCAGCAATCCGATCCGATCATCAACGCCCGCGCACAC 943237  
QY 281 ThrSerSerPheHisValIleGlyIlePheAspLysValHisPheGlnGly----- 298  
DB 943236 GCGTTCGCGATCGCGTGGCGCGGCGATTCGATGACGCTACCAACCGACGCTTACCA 943177  
QY 299 -----GlyLysGlyGluAsnHisAsnIleGln 307  
DB 943176 GTGATTCACCGAAGTCGACGCTGTGATGCGGATGCGGCGCAACGCTACGACGTCATG 943117  
QY 308 ThrThrLeuIleProAlaGly-----AlaAlaIleThrGluPheLysVal 323  
DB 943116 GTGACC-----GCCGTGGCGGCTTTCCCTGCGTGCATCGGAGGCAAG--- 943066  
QY 324 AspValProGlyAspTyrValLeuValAspHisAlaIlePheArgLysPheAsnLysGly 343  
DB 943065 -----AACGCTGGCGCGTCCGCTGCTTAC 943036  
QY 344 AlaLeuGlyIleLeuLysValGluGlyGluAsnHisGluIleTyrSerHisLysGln 363  
DB 943035 GCGCGCGCG-----ACCGCA 943021

[illegible]

```

QY 299 -----GlyylsglyluasnHisasnIleGln 307
Db 943286 GTGATTCACACCGAAGTCAGACGTCTGTGATCGCATGGCGAACCCTACAGATCATG 943227
QY 308 ThrThreuleProAlaGlyly-----AlaAlaIethrGluPheVal 323
Db 943226 GTGACC-----GCCGCTGGGGCGCTTCCCTCGTGCACCTCGCGAAGCAG--- 943176
QY 324 AspValProGluAspTyrValValLeuValAspHisAlaIethrPheArgAlaPheAsnGly 343
Db 943175 -----AACGCGCTGGCGCGCTGCTGCTCTTACC 943146
QY 344 AlaLeuGlyIleLeuValGluGlyGluGluAsnHisGluIleTyrSerHisIleGln 363
Db 943145 GGGCGCGCG-----AGCCCA 943131
QY 364 ThrAspAlaValTyrLeuProGluGlyAlaProGlnAlaIleAspThrGlnGluAlaPro 383
Db 943130 CCGGACCCGAGTTTCGGCGCGAAGTCAACTGCGAGTGGGTACCCGTGAATGTTTC 943071
QY 384 LysThrProAlaProAlaAsnLeuGlnGlnIleLysAlaGlyLysAlaThrTyrAsp 403
Db 943070 ACCGCGCCAAAGACTGCGCAACTG----- 943047
QY 404 SerAsnCyAlaAlaCyHisGlnProAspGlyLysGlyValProAsnAlaPheProPro 423
Db 943046 -----GGCGCGCCGAGACCACCCAGCAGCTCCG 943017
QY 424 LeuAlaAsnSerAspTyrLeuAsnAlaAspHisAlaArgAlaAlaSerIleValAlaAsn 443
Db 943016 -----GTCACCTTGGGC 943005
QY 444 GlyLeuSerGlyLysIle-----ThrValAsnGlyAsnGlnTyrGluSerValMetPro 461
Db 943004 GGACCATGGCCAAAGTACGACTGCACATCAACGGGAAACCTTACACAGCACAATCA 942945

RESULT 31
AAK57912
ID AAK57912 standard; DNA; 5187 BP.
AC AAK57912;
XX
DT 15-JUL-1999. (first entry)
XX
DE G. oxydans D-sorbitol dehydrogenase coding sequence.
XX
KW D-sorbitol dehydrogenase; L-sorbose; 2-keto-L-gulonic acid; precursor;
KM L-ascorbic acid production; ss.
OS Gluconobacter oxydans.
XX
PN MO9920763-A1.
XX
PD 29-APR-1999.
XX
PF 13-OCT-1998; 98MO-JP04612.
XX
PR 17-OCT-1997; 97JP-0285280.
XX
PA (FUJII) FUJISAWA PHARM CO LTD.
XX
PI Iehi Y, Noguchi Y, Saito Y, Soeda S, Yoshikawa K;
XX
DR WPI; 1999-302741/25.
XX
PT Game group for D-sorbitol dehydrogenase, useful for simple
PT large-scale production of L-sorbose or 2-keto-L-gulonic acid as
PT precursor for L-ascorbic acid
PS Claim 21; Page 60-62; 83pp; Japanese.
XX
CC This sequence encodes the D-sorbitol dehydrogenase of the

```

```

CC invention. Cells transformed with a vector containing DNA encoding
CC the dehydrogenase can be used to produce L-sorbose or 2-keto-L-gulonic
CC acid precursor for simple large-scale L-ascorbic acid production.
XX
SQ Sequence 5187 BP; 1251 A; 1390 C; 1297 G; 1249 T; 0 other;

Alignment Scores:
Pred. No.: 0.00128 Length: 5187
Score: 155.00 Matches: 133
Percent Similarity: 33.17% Conservative: 68
Best Local Similarity: 21.95% Mismatches: 207
Query Match: 5.90% Indels: 199
DB: Gaps: 29

US-10-088-045-2 (1-502) x AAK57912 (1-5187)

QY 15 LeuSerAlaLeuMetLeuSerGlyCySerAsnGlnAlaAspLys----- 29
Db 2667 CTGATGGCTGTGATG--AGTGGTACTGATCTTCAATGACAAAGGTTACTTCCGCGAG 2723
QY 30 -----AlaAlaGlnProLysSerSerThrValAspAla 40
Db 2724 TGGCATCTCGGGGACAGATCATGAGAAACAGCCCAACCAATTCGTCGTTGACGGT 2783
QY 41 AlaAlaLysThrAlaAsnAlaAspAsn-AlaAlaSerGlnGluHisGlnGlyLysLeuPro 60
Db 2784 GAGTGGCGGACCATGATGACGAAACCTGTTGTCGATGACGAGCGCTTCTCTTGG 2843
QY 60 ValIleAspAlaIleValThrHisAlaProGluValProProProValAspArgAspHis 80
Db 2844 GTCGCTACAGCAGATA-----TCACCTGACCAATTTGGCGCC----- 2880
QY 80 sProAlaLysValValValLysMetGluThrValGluLysValMetArgLeuAlaAspG1 100
Db 2880 ----- 2880
QY 100 ValGluTyrGlnPheThrPheGlyGlnValPro---GlyGlnMetIleArgVal 119
Db 2881 -----TGGCGCTTCGGCTTGCAGATCCCTGAAAGAGAGATGCTTCAT 2924
QY 119 LArgGlyLysAspThrIleGluValGlnPheSerAsnHisProAspSerLysMetProHis 139
Db 2925 GCGTGAGGGGAATGAACCCGGAATACGCCCTTTCTGCGACAGCTCC----- 2973
QY 139 sAsnValAspPheHisAlaIaIaThrGlyProGlyGlyValAlaGluAlaSerPheThrAl 159
Db 2974 -----ATAGCTTCGGGTGCTGCTTTCGGCGCGAG--TCAGCGAGGCG 3014
QY 159 aProGlyHisThrSerThrPheSerPheLysAlaLeuGlnProGlyLeuTyrVal----- 177
Db 3015 AGAGGATCAGGACCACTATCAGC-----CGAGCGCCCTATCTGGGCTAC 3059
QY 178 -----TyrHisCysAlaValAlaProValGlyMetHisIleAlaAsnGlyMe 193
Db 3060 AGCAGGCACTGGCTTCCCTGCATACGAAACAGGTGGGCTCCCTTTCGGGCGC----- 3114
QY 193 tTyrGlyLeuIleLeuValGluProLysGlyLysLeuProLysValAspLysGlyTyrTyr 213
Db 3115 -----GGCTTGTATTCGCTCCCAATGCGGGAGTGT- CGCGTC-----CAACATTA 3163
QY 213 rValMetGlnGlyAspPheTyrThrLys-GlyLysTyrGlyGlnGlnGlyLeuGlnProP 233
Db 3164 CACCCGATCCGGA-----TACGGGAATTGGCAAAATACACCGAAGAGAGATTGCCAAC- 3216
QY 233 heAspMetGlyLysAlaIleArgGluAsp----- 242
Db 3217 --GCTTTCGCAAGGAGTTCGACAGGACCGAGCTCATCTTACCGGCATGCTTACA 3274
QY 243 -----AlaGluTyrValValPheAsnGly 251
Db 3275 CGGCTATTTCGAGATTGGGATTCGAGATCCAGCATCCAGCATTTGATGTTACTCATGCGATG 3334
QY 251 ervaIaGlyAlaLeuThrGlyLysAla-----LeuL 262

```

```

Db 3335 GCGTGGCCCCCTGGCGGAGCAAAATCCGAGAGCTGAATTCCTCAATATCC 3394
Qy 262 YSAlaIyValGlyIuThrValArgIeuPheValGly-----AengIyGlyP 278
Db 3395 GCGCATGATGATGACAGCTGGAATCTCTGTTCAGAGACCTCCGCCGGAAGGGTATG 3454
Qy 278 rOaenIeuThrSerPhe-----HisValIleGlyGluIlePheAspIysValH 295
Db 3455 CTCAGACCTATTCGACAAATCGAAGAGGCCATATCTGSCAATGCGCTTG 3504
Qy 295 IephGlyGlyGlyIyGlyGlyIuAsnHisAsnIleGlnThrIle----- 311
Db 3505 -----GGACATTCGCGAAGCTGTGATACACACGCAATTCCTGATGGCGAAGCA 3556
Qy 312 -----P 312
Db 3557 GCAGCAGTCCCTATCTGGCGGAGACCGCGCTGCTGATGCTCCCAATCAGC 3616
Qy 312 rOAlaGlyGlyAlaAlaIleThrGluPheIyValAspValProGlyAspIyValIeu 332
Db 3617 CGAGCATGAATGCGCGGATCGCGCATGTGAGCGAAGCATGTGTTCAATCTCGCTA 3676
Qy 332 aIAspHisAlaIlePheArgAlaPheAsnIyGlyAlaIeuGly----- 346
Db 3677 CAGGCTCCGCGCAGAGCGCTCTCAGCGCGCATGATGCGAAGCTGTTGAACATA 3736
Qy 347 --IleIeuIyValGlyIuGlyIuAsnHisGlnIleIySerHisIyGlnIleAsp 366
Db 3737 GCTTTAGCAAGCTGACAGACGAGATCTCCACGCGATC-----G 3775
Qy 366 IAspIyIeuProGlyIyAlaProGlnAlaIleAspThrGln----- 380
Db 3776 CCGCTATATCCGACAG---ATCCGAAGATGAGAGACGCAACCAAAACGCGCTG 3832
Qy 381 -----GluAlaProIyThrProAlaP 388
Db 3833 ACCGCTTCGGGTTGCCGTCAGCCCATCTGATCTGCAGAACCAAACTGTGCTG 3892
Qy 388 rOAlaAsnIeuGlnIuGlnIleIyValGlyIyAlaIleThrIyAspSerAsnCyAla 408
Db 3893 AAGATGACCTGTTCCGATG-----GACGGGAGAGATCTACCTCAACTGTGCA 3946
Qy 408 IAspHisGlnProAspGlyIyGlyIyValProAsnAlaPheProIeuAlaAsn 428
Db 3947 CCTGCGCATGACTGATGTCAGAGAGCGCGGACCTTCAAGCTCTCTGCTTCA 4006
Qy 428 sPTyIyIeu-----AsnAlaAspHisAlaArgAlaAlaSerIleValAlaAsn 444
Db 4007 ATGCAGTAGTCGTCACCGGAGCTGACAACTGATCATGAGCC-----ATTGCAACG 4060
Qy 444 IyIeuSerGlyIyIleThrValaAsnIyAsnIyGlnIyIuSerIleAspValIle 464
Db 4061 GCGTTGATGCC-----ACGACGAATGGT-----CATCAGCTTGTGATGCGGGTTTCG 4108
Qy 464 Ia-----LeuSerAspGlnIleAlaAsnValIleThrIyT 477
Db 4109 GCCCAGCTTCGATGATACAGCGCTCAGCGATACGAGATGCGGAATCTCACCACTATG 4168
Qy 477 hIleAsnSerPheGlyIyAsnIyGlyIyGlnIeuSerAlaAspValAlaIyAla 497
Db 4169 TCTCCGAGCATTTTGGAGTGGCATCATGATCATCAGCTCAGACCTAAAGTGGCTC 4228
Qy 497 ySlyIyThIyAspPro 501
Db 4229 GTGAAGCGCGGCT 4242

```

RESULT 32

```

ABT14589
ID ABT14589 standard: DNA, 1125 BP.
XX ABT14589;
AC
XX

```

```

DT 27-FEB-2003 (first entry)
XX
DE Pseudomonas aeruginosa biofilm formation-related gene #1.
KM Gene; ds; biofilm formation modulation; biofilm-associated disease;
KM cystic fibrosis; AIDS; middle ear infection; acne; periodontal disease;
KM catheter-associated infection; medical device-associated infection.
XX
OS Pseudomonas aeruginosa.
PN W0200285295-A2.
XX
PD 31-OCT-2002.
XX
PF 19-APR-2002; 2002MO-US12532.
XX
PR 20-APR-2001; 2001US-285190P.
PR 24-OCT-2001; 2001US-344142P.
XX
PA (IOWA ) UNIV IOWA RES FOUND.
PA (HARD ) HARVARD COLLEGE.
XX
PI Whiteley M, Banger MG, Lory S, Greenberg EP;
XX WPI; 2003-075601/07.
XX DR P-PSDB; ABT18767.
XX
PT Identifying compound capable of modulating biofilm formation by
PT bacterial/bacterial antibiotic resistance, useful for treatment of
PT biofilm associated disease -
XX
PS Claim 1; Page 75-76; 154p; English.
XX
CC The invention comprises a method for identifying a compound capable of
CC modulating biofilm formation by bacteria. The method of the invention is
CC useful for identifying a compound capable of modulating biofilm formation
CC by bacteria or modulating bacterial antibiotic resistance. The method of
CC the invention is also useful for diagnosing and treating a subject
CC (especially an immunocompromised human) that is afflicted with a biofilm-
CC associated disease or disorder, such as: cystic fibrosis; AIDS; middle
CC ear infections; acne; periodontal disease; catheter-associated
CC infections; and medical device-associated infections. The present DNA
CC sequence represents a gene that is used in the invention.
XX
SQ Sequence 1125 BP; 221 A; 362 C; 361 G; 181 T; 0 other;

```

Alignment Scores:

Pred. No.:	Length:	Matches:	Score:
152-50	1125	87	152.50
Percent Similarity:			31.83%
Best Local Similarity:			21.80%
Query Match:			5.81%
Indels:		162	
Gaps:		18	

```

US-10-088-045-2 (1-502) x ABT14589 (1-1125)
Qy 132 HisProAspSerIyMetProHisAsnValAspPheHisAlaIleThrGlyProGly 151
Db 318 CACCTTAT-----CCACATCTAGACACACTTCCGAGCGGAGCTGAGCGTGA 368
Qy 152 -----GlyAlaGlnAlaSerPheThrAlaProGlyHisThrSerThrPheSerPhe 168
Db 369 CACCGGCTACAGAGAGTGGCGATCAAGTACCTGG----- 407
Qy 169 IyAlaIeuGlnProGlyIeuIyValIyThiGlyAlaValAlaProValGlyMetHis 188
Db 408 -----CCAGAGCTCGAGTACTTACAGAACCTGCGCACCGCGAGATCA--- 452
Qy 189 IleAlaAsnIyMetIyGlyIleIeuValGlnProIyGlyIyIeuPro-IySva 208
Db 453 -----GATCCACAA-----CCGCGAGGC 470
Qy 208 IAspIyGlyIyIyIyValMetGlnIyAspPheIyThrIyGlyIyIyGlyIyGlnI 228

```

```

Db      471 GAAGAGCAGCATTAAGCTGAGGTGAC----- 501
Qy      228 nglyleuGlnProPheaspMetGluValaIleArgGluaspIaGluTyrValaPh 248
Db      501 ----- 501
Qy      248 eaanglySerValGluValaLeuThrGluGluaspIaLeuValaGluTyr 268
Db      502 ----- 530
Qy      268 rValaGluPheValaGluanglyGluProaspLeuThrSerSerPheHisValaIleGlu 288
Db      531 GGTCGGCTTCCTGATC-----ACCTCCAGC----- 555
Qy      288 yGluIlePheAspLysValaHisPheGluGluGluValaGluValaHisValaIleGln 308
Db      556 -----GACGTATCCAT-----TC 569
Qy      308 rThrLeuIleProIaGluGluVala-----AlaIleThrGluPhe----- 321
Db      570 CTGGTGGTGCCTGGCTTCGGCGTCAGCGCGAGCCATCCCGGCTTCGTCAGAGGC 629
Qy      322 -----LysValaAspValaProGluAspTyr----- 329
Db      630 CTGAGCCAGAGTGCAGCGAGCCGCGCATCTATCCGCGCAGTCGCGCAGCTGTGCGCAA 689
Qy      330 -----ValleuValaAspHisAlaIlePheArgAlaPheAsnLysGlu 343
Db      690 GGAACACGCTTACGCGCATCGGTGTCGAC----- 720
Qy      343 yAlaLeuGluYlleuLysValaGluGluGluValaAsnHisGluIleTyrSerHisLysGlu 363
Db      721 -----GTCAAGCCCAAGCGCGAGTTCGACAGTGGTGGCCAAAGCCAAAGCA 767
Qy      363 nThrAspAlaValaTyrLeuProGluGluYalaProGlnAlaIleAspThrGlnGluAlaPr 383
Db      768 A--GAGCGCGCGAGGTCAGAGAA-----CTGACCAAGCAAGAGTGAGAC 809
Qy      383 oLysThrProAlaProIaAsnLeuGlnGlnIleLysAlaGluYalaThrTyrAs 403
Db      810 CAAG-----GAAGAGTTGGTTCGCGCGCGCAAGAGTCTTCCA 848
Qy      403 pSerAsnCyAlaAlaCysHisGlnProAspGluYValaProaspIaPheProPr 423
Db      849 CACCATTCGCGCGCTGCACAGCGCGCAAGCGCAGGCGATGTCCTCCGCGC 908
Qy      423 oLeuAlaAsnSerAspTyrLeuAsnAlaAspHisAlaArgAlaAlaSerIleValaAs 443
Db      909 GCTGAAGGTTGCAAGATCGTCACCGGCGCCAGAGACACCACTGGAAGTGTCTTCA 968
Qy      443 nglyLeuSerGly-----LysIleThrValaanglyAsnGlnTyrGluSerValaMetProAl 462
Db      969 CGGCGTGGCCGCGAGCCGCGCATGGCGGCTTCGCGCAACAG----- 1008
Qy      462 aIleAlaLeuSerAspGlnGlnIleAlaAsnValaIleThrTyrThrLeuAsnSerPheG 482
Db      1009 -----CTCAAGAGGTCAACCTGCGCGGTATCACTTCAAGCGCAAGCGCTGGG 1061
Qy      482 yAsnLysGluYlleuLysSerAlaAspValaIleLysAlaLysThrTyrHis 500
Db      1062 CAACGACGATGGCGACATGTGTACCCGAAAGACGTGTGCGCTTCAAGAGCAAA 1116

```

## RESULT 33

AAQ20384 standard; DNA; 1404 BP.

AAQ20384;

16-APR-1992 (first entry)

ADH complex structural gene (44,000 mol. wt. protein).

```

KW      Alcohol dehydrogenase; acetic acid; fermentation; ss.
XX
OS      Acetobacter alioacetigenes IM-24.
XX
PN      JF03266988-A.
XX
PD      27-NOV-1991.
XX
PF      26-MAR-1990; 90JP-0073440.
XX
PR      26-FEB-1990; 90JP-0042301.
XX
PR      26-MAR-1990; 90JP-0073440.
XX
PA      (NAKA-) NAKANO SUTEN KK.
XX
DR      WPI; 1992-019325/03.
XX
DR      P-PSDB; AAR20193.
XX
PT      Alcohol dehydrogenase complex structural gene - used in plasmid
PT      and enhancing efficiency of acetic acid fermentation for
PT      transformed acetic acid bacteria
XX
PS      Disclosure; Fig 4(1-3); 21pp; Japanese.
XX
CC      The gene encodes a protein of mol. wt. ca. 44,000. Acetobacter
CC      transformed with the sequence can enhance the efficiency of acetic
CC      acid fermentation. The ADH complex can be easily extracted from the
CC      bacteria and purified and it can be used for the determination of an
CC      alcohol.
CC      See also AAQ20383-84, and -86-88.
XX
SQ      Sequence 1404 BP; 253 A; 448 C; 437 G; 266 T; 0 other;

```

## Alignment Scores:

Pred. No.:	0.000415	Length:	1404
Score:	152.00	Matches:	108
Percent Similarity:	38.02%	Conservative:	57
Best Local Similarity:	24.88%	Mismatches:	165
Query Match:	5.79%	Indels:	104
DB:	13	Gaps:	23

```

US-10-088-045-2 (1-502) x AAQ20384 (1-1404)
Qy      113 ProGlyValMetIleArgValaGy-----GluGluAspThrIleGluValaGln 128
Db      156 CAAATGACAGCTGTACAGCGCGGCTTGAATCAAGAGCCGATCGTACGATCTCTC 215
Qy      129 PheSerAsnHis-----ProAsp----- 134
Db      216 CACGAACATCACACCGAGCCGACCTACGCTATCGGTACACCTTCGCGCAATTGCA 275
Qy      135 ---SerLysMetPro-----HisAsnValaPheHisAlaAla 146
Db      276 CGAAGCGTGGCGCATGTATTCGCAAGACGGTTCACACCTGTATCGGCGCAATGCGGTA 335
Qy      147 Thr-----GlyProGluYlleuGlnIleGluValaIleSerPheThrAla 159
Db      336 TCCCTCTTTCGCGCATGACGAAGAGACATGCAAGCGCTGTATCGTACTTCATGCA 395
Qy      160 ProGluHisThrSerThrPheSerPheLysAlaLeuGlnProGluLeuTyrValaTyrHis 179
Db      396 TGGGGTGAAGCCGGTGGCGGAGCGGACAGCAAGCGGAGCATCTCCGCGCTTGTGCAT 455
Qy      180 CysAlaValaIaProValaGlyMetHisIleAlaAsnGlyMetTyrGluLeuIleVala 199
Db      456 GCGCTGCGCGCTGG-----CATCTGCGCATGAT-----GTTTCGCGC 494
Qy      200 -GluProLysGluGluYlleuProLysValaAspLysGluTyrTyrValaMetGlnIleAsp 219
Db      495 TTGCGCGAAGAGCTTACAGCGCGCGGCGGACGAGCATCTGGAATGCAAGCTGCGCATTA 554
Qy      219 eTyrThrLysGluYlleuTyrGluGlnGluLeu-----GlnProPheAsp 235

```

```

Db 555 TCTGTTACGGGCCCC---GGGCATTCGGGTCGTCATACACCCGGGCTTCGGCAT 611
Qy 235 T---GluValAlaIleArgValAspAlaGluTyrValValPheAsnGlySerVal-GlyA 254
Db 612 GCAGGAAAGGCGCTGCAGCGCTCCGGTCTCTGACTTCTGTCTCGGTGGCGACCGAT 671
Qy 254 ILeuThrGlyGlyValAsnAlaLeuValAlaGlyGlyValGlyValArgLeuPheValG 274
Db 672 GCACAACTGGGTCCGGCGGACGTCGGCA-----ACGATCTGTGCTGGTCT 719
Qy 274 LyuAsnGlyGlyProAsnLeuThrSer--SerPheHisValIleGlyGlyLeuPheAsp 293
Db 720 GGGCGCGCTGTCTCCGAGATGACATCTACCTCTTAAGTCCGCGCGCTGACCACTC 779
Qy 293 ValHisPheGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 313
Db 780 CGCGGCTGTC---GTTGGCATGGCGCAT----- 804
Qy 313 AGlyGlyAlaAlaIleThrGlyPheValAspValProGlyAspTyrValLeuValAs 333
Db 805 -----GTGGTGGCATGGAGACCCCACTACTTCAACGATGACGACCTG----- 846
Qy 333 PHisAlaIlePheArgAlaPheAsnGlyGlyAlaLeuGlyTyrLeuValGlyGlyGly 353
Db 847 -CACGCCCATCGCGAG---TACCTGAAGAGCTCCCGCGCTCCGCTCACAGGCGCA 902
Qy 353 uGluAsnHisGlyLeuTyrSerHisGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 373
Db 903 C-----TACACCTAGCATCCGCTCCACCGGACGACCTGCTGGCTGGCT-- 945
Qy 373 AProGlnAlaIleAspThrGlnGlyAlaProGlySerProAlaProAlaAsnLeuGlnG 393
Db 946 -----AATACCGCGACGCTTCG----- 963
Qy 393 uGlnIleLeuAlaGlyValAlaTyrTyrAspSerAsnGlyAlaAlaCysHisGlnProAs 413
Db 964 -----GGTCTGATAGTATGTGAAGAAATGCGCCATCTGTCAACCTTAACGA 1010
Qy 413 PGIlyGlyGlyValProAsnAlaPheProProLeuAlaAsnSerAspTyrLeuAsnAlaAs 433
Db 1011 CGGTGTGGCGTGGCGCGCATGTTCCGCGCTGGCGCAACCGGTTGTCGTACCGCA 1070
Qy 433 PHisAlaArgAla---AlaSerIleValAlaAsnGlyLeuSerGlyTyrHisIleThrValAs 452
Db 1071 GAACCCGACCTCGCTGTGAACGATGGCGCATGCT-----GGCGTGGCGCGCGAG 1124
Qy 452 nGlyAsnGlnTyrGlySerVal---MetProAlaIle-----AlaLeuSerAspGlnG 469
Db 1125 CAACCTGGGCAACCTCCGACAGTGCAGATGCGGGTTACAGCAAGATGCTGTCGCCCGCA 1184
Qy 469 nIleAlaAsnValIleThrTyrThrLeuAsnSerPheGlyAsnGlyGly---GlyGlnLe 488
Db 1185 GATTGCTGATGTGCTCACTTCATCCGACACCTGGGCAACAGGCGCGCGACCGT 1244
Qy 488 uSerAlaAspAspValAlaLeuValAlaGlyGlyGly 499
Db 1245 TACGGCTGGCGATGTTACCAAGCTGGCGACGACG 1278

```

## RESULT 34

AAQ13581 standard; DNA; 1404 BP.

```

XX AAQ13581;
XX
XX 25-MAR-2003 (updated)
XX 09-DEC-1991 (first entry)
XX
XX A.altoacetigenes membrane-bound ADH 44kD sub-unit.
XX
XX alcohol dehydrogenase complex; carboxylic acid production;
XX cytochrome c; ss.
XX
XX Acetobacter altoacetigenes.
OS

```

```

XX XX EP448969-A.
XX PN 02-OCT-1991.
XX PD
XX XX 26-FEB-1991; 91EP-0102793.
XX PF
XX PR 26-MAR-1990; 90JP-0073440.
XX PR 26-FEB-1990; 90JP-0042391.
XX
XX (NAKA-) NAKANO VINEGAR CO LTD.
XX (NAKA-) NAKANO VINEGAR CO LTD.
XX PI Tamaki T, Takemura H, Tayama K, Fukaya M, Okumura H;
XX PI Kawamura Y;
XX DR WPI; 1991-289462/40.
XX DR P-PSDB; AAR13994.
XX
XX Gene for membrane-bound alcohol dehydrogenase complex - obtd.
XX PT from Acetobacter altoacetigenes, used for prodn. of enzyme for
XX PT converting alcohol to acid
XX
XX Claim 4; Page 23-24 and Fig 4; 36pp; English.
XX
XX Total DNA was prepared from A.altoacetigenes MH-24, digested with
XX CC PstI and SmaI and ligated to pSci-SmaI cleaved pUC18. The ligated
XX CC mixture was used to transform E.coli JM109. Probes were designed
XX CC based on the N-terminal amino acid sequence of the 72kD subunit of
XX CC the ADH complex isolated from A. altoacetigenes (see AAQ13582-Q13584).
XX CC Fragments encoding the 44kD (cytochrome c) ADH subunit were
XX CC identified using anti-44kD subunit antibodies in a Western blot.
XX CC The gene was sequenced. Various experiments suggested that the
XX CC coding sequence is located immediately downstream of the gene
XX CC encoding the 72kD sub-unit. See also AAQ13580.
XX CC (updated on 25-MAR-2003 to correct PA field.)
XX
XX SQ Sequence 1404 BP; 254 A; 44% G; 436 G; 266 T; 0 other;

```

## Alignment Scores:

```

Pred. No.: 0.00628 Length: 1404
Score: 150.00 Matches: 108
Percent Similarity: 37.79% Conservative: 56
Best Local Similarity: 24.88% Mismatches: 166
Query Match: 5.71% Indels: 104
DB: 12 Gaps: 23

```

US-10-088-045-2 (1-502) x AAQ13581 (1-1404)

```

Qy 113 ProGlyGlnMetIleArgValArg-----GluGlyAspThrIleGluValGln 128
Db 156 CCATGACAGCTCTTACGACAGCGCGCTTGAATCAAGAGCCGATCGTACGATCTATC 215
Qy 129 PheSerAsnHis-----ProAsp----- 134
Db 216 CACGAACATCAACACCGACCGACCTACGATCGCTACACCTTCCGCAATTCGA 275
Qy 135 ---SerIlyMetPro-----HisAsnValAspPheHisAlaAla 146
Db 276 CGAAGCGCTGGCGCATGATTCGCAAGACGGTTCACGCTTATCCGCGACCGGTA 335
Qy 147 Thr-----GlyProGlyGlyValGlyValGlyValGlyValGlyValGlyVal 159
Db 336 TCCCTCTCTTCCGCGATGACGAAGAGACATGACAGGCGCTGATCCGATTCATGCA 395
Qy 160 ProGlyHisThrSerThrPheSerPheValAlaLeuGlnProGlyLeuTyrValTyrHis 179
Db 396 TGGGATGAGCGCGTGGCGGACCGGACCAAGACCGGACATCTCGTGGCGCTTGCAT 455
Qy 180 CysAlaValAlaProValGlyMetHisIleAlaAsnGlyMetTyrGlyLeuIleLeuVal 199
Db 456 GCGCTGGCGCGCTGGG-----CATCTGGCGCATGAT-----GTTCTCGCC 494

```



```

RESULT 36
AAH66018
ID AAH66018 standard; DNA, 1533 BP.
XX
XX
XX AAH66018;
XX
XX
XX 26-SEP-2001 (first entry)
XX
XX C glutamicum coding sequence fragment SEQ ID NO: 1053.
XX
XX Colyiform bacterium; amino acid synthesis; vitamin; saccharide;
XX organic acid synthesis; ds.
XX
XX Corynebacterium glutamicum.
XX
XX EP1108790-A2.
XX
XX 20-JUN-2001.
XX
XX 18-DEC-2000; 2000EP-0127688.
XX
XX 16-DEC-1999; 99JP-0377484.
XX
XX 07-APR-2000; 2000JP-0159162.
XX
XX 03-AUG-2000; 2000JP-0280988.
XX
XX (KYOW ) KYOWA HAKKO KOSYO KK.
XX
XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
XX Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX WPI: 2001-376931/40.
XX
XX P-PSDB; AAG90799.
XX
XX Novel polynucleotides derived from Coryneform bacteria, for identifying
XX mutation point of a gene, measuring expression of a gene, analysing
XX expression profile or pattern of a gene and identifying homologous gene
XX
XX
XX
XX
XX Claim 8; SEQ ID NO: 1053; 246bp + Sequence listing; English.
XX
XX
XX The present invention provides a number of nucleotide and protein
XX sequences from the Coryneform bacterium Corynebacterium glutamicum. These
XX are useful for identifying the mutation point of a gene derived from a
XX mutant of coryneform bacterium, measuring expression amount and
XX analysing the expression profile or expression pattern of a gene derived
XX from Coryneform bacterium, and identifying a homologue of a gene derived
XX from coryneform bacterium. Coryneform bacteria are useful for producing
XX amino acids, nucleic acids, vitamins, saccharides and organic acids,
XX particularly L-lysine. The present sequence is a nucleic acid described
XX in the exemplification of the invention.
XX
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from the
XX European Patent Office.
XX
XX
XX Sequence 1533 BP; 321 A; 415 C; 458 G; 339 T; 0 other;
XX
XX
XX Alignment Scores:
XX
XX Pred. NO.: 0.00372 Length: 1533
XX Score: 142.00 Matches: 99
XX Percent Similarity: 33.72% Conserved: 47
XX Best Local Similarity: 22.86% Mismatches: 178
XX Query Match: 5.41% Indels: 109
XX DB: Gaps: 20
XX
XX
XX US-10-088-045-2 (1-502) x AAH66018 (1-1533)
XX
XX
XX 29 LyeAlaAlaGlnProLysSerThrValAlaPAlaAlaLysThrAlaAsnAlaAsp 48
XX ||||| :|||:|||||
XX 31 AAAAGGCGCCGGGCGTGTGGCGAACAAGGTGTGGCGCCAGGTGTCGTGGCGGTTC 90
XX ||||| :|||:|||||
XX 49 AsnAlaAlaSerGlnGlnHISGlnGlyGlu-----LeuProValIleAspAlaIle 65
XX :|||:|||||
XX 91 TCGATGATGATGTGCGTGTATGGGAGAGACCGCGGACCTGTCCTCAATT----- 138

```

OY	66	ValThiHisAlaProGluValProProProValAsp-----AcAspHisProAlaLys	83
Db	139	-----CCACCGACGAAATTAGGTACGGCGAGGAACTTACG	174
OY	84	ValValValLysMetGluThrValGluLysValMetArgLeuAlaAspGlyValGluTr	103
Db	175	GTGCACCTTGGCCCTCGAGAGCTCAGACCTGGGAGAGATGATTTGGCCGAGATTCAACG	234
OY	104	GlnHetPrrThrPheGluGlyGlnValProGlyGlnMetIleArgValArgGlnGlyAsp	123
Db	235	AAAGCGTGGGTTTCAATTGGCACTCATTTGGGCGCGACGTTGGTGTAAGAAAGGTAT	294
OY	124	ThrIleGluValGlnPheSerAsnHisProAspSer-----	135
Db	295	GACGTCCACGTTGATGTGATTAACAATTGGATGAATGACCACTGTGCATCGATTCG	354
OY	136	---LysMetProHisAsnValAspPhe-----HisAlaAlaThrGlyProGlyGly	152
Db	355	ATGAAAGTTGGCCGGATGTTCTGATGGTGGTCCGACCTACCGATGGGCGCTGGGCAAGC	414
OY	153	AlaGluAlaSerPheThrAlaProGlyHisSerSerThrPheSerPheLysAlaLeuGln	172
Db	415	TGCGTACCAACGTGACTGTGGCCAAATGATGCACCACTTTGTGTGTACACCCGCACT	474
OY	173	ProGlyLeuValTyThrHisCysAlaValAlaProValGlyMetHisIleAlaAsnGly	192
Db	475	CATGGCGCTG-----ACAGGTTTGCATGCGCTACCGTGGT	507
OY	193	MetTyrglyLeuIleLeuValGluProLysGlnGlyLeuProLysValAsp-----Lys	210
Db	508	TTCGGCGGGATGATCATTTGTGGAA--GATGAACAACAGACAAAGCTGGATCTGCACGC	564
OY	211	GluTygTygVal-----MetGlnLysAspPheTyThrLysGly	223
Db	565	GAGTACGATGGGACGATATTCGCGTGTTTATGATCACCGCTCTTGAGAAAGCGT	624
OY	224	LysTyrglyGluGlnGlyLeuGlnProPheAspMetGluLysAlaIleArgGluAspAla	243
Db	625	TCCCTTGAATAGGAAGACCTCCG---GATCTTGAGCGTTGGCGGATACCCCACT	678
OY	244	GluTygValValPheAsnGlySerValGlyValAlaLeuThrGlyGluAsnAlaLeuLysAla	263
Db	679	GCCAAATGGCATTTACCAATGCCGCACTTGATGTGCCACACGCGCGGTTGCGTCCGCGT	738
OY	264	LysValGlyLeuThrValArgLeuPheValGlyAsnGlyGlyProAsnLeu-----	280
Db	739	CTCAACGGCTCCAAATATGCGGTTCTAT-----AATCTGGCGTTTCA	780
OY	281	---ThiSerSerPheHisValIleGly-----GluIlePheAsp	292
Db	781	GACACGGGCACTTCCAAATGATTTGGCCAGCGATTCGCGTTGCTGATGAACCTCAAGAC	840
OY	293	LysValHisPheGlnGlyGlyLysGlyGluAsnHisAsnIleGlnThrThrLeuIlePro	312
Db	841	CGCACACCTTGGCTATTTGGCGCCAGCGAGCGGAGAAATCGTCGAGACTAGAGCC	900
OY	313	AlaLeuGlyAlaAlaIleThr-----GluPheLysValAspValPro	326
Db	901	-----GCGAGAGCGTCACTTGGATCTGTAGAGTTTGAAGCAACTACGCGGCTCCT	954
OY	327	GlyAspTygValValLeuValAspHisAlaIlePheArgAlaPheAsnLysGlyAlaLeuGly	346
Db	955	GATATATAGTTTCGCGCCGATTTTGGCATGTGCATCTTCCTCCAG-----	999
OY	347	IleLeuLysValGluGlyGluAsnHisGluIleTygSerHisLysGlnThrAspAla	366
Db	1000	CTGCTTCAACATCAACGCGCTCCGAT-----GATGCT	1032
OY	367	ValTygLeuPro-----GluGlyAlaProGlnAlaIleAsp	378
Db	1033	GCGCAACACCTGTTTCCGCGCGCTGCTGTGTGAATTCACCGAATGACGTATGATGAT	1092







Db	101806	CATGCGCTG-----	ACAGGTTTGATGGGTACCGGCT	101838
Qy	193	MecTYrGlyLeuIleuValGluProLysGluGlyLeuProLysValAsp-----Lys	210	
Db	101839	TTGGCGGGGATCATCTGTGGAA---GATGAAGCAACAGACGCTGGATCTGCCACGC		
Qy	211	GluTYrTYrVal-----MetGlnGlyAspPheTYrThrLysGly	223	
Db	101896	GAGTACGGTGTGGACGATATTCCGCTGTTTATGATGATCACCGCTTCTTAGAAGACGGT	101955	
Qy	224	LysTYrGlyGluGlnGlyLeuGlnProPheAspMetCyluValAlaIleArgGluAspAla	243	
Db	101956	TCCCTGATGAGGAAGACCTCCCC-----GACTTGGCGGTGTGGGCGCATACCACCAC	102009	
Qy	244	GluTYrValValPheAsnGlySerValGlyAlaLeuThrGlyGluAsnAlaLeuLysAla	263	
Db	102010	GCCAAATGACATTCACCAATCCGACCTTGGATGCCACACGCGCGGGTTCGGTTCGCCGTG	102069	
Qy	264	LysValGlyGluThrValArgLeuPheValGlyAsnGlyGlyProAsnLeu-----	280	
Db	102070	CTCAACGGCTCCAAATATGCGGTTCTAT-----AACTTGGCGTTTCA	102111	
Qy	281	---ThrsSerPheHisValIleGly-----GluIlePheAsp	292	
Db	102112	GACACGCCACCTTCCAACTGATTCGACACGATCCCGTTTCTGTGATGACCTCAAGAC	102171	
Qy	293	LysValHisPheGluGlyGlyLysGlyGluAsnHisAsnIleGlnThrThrLeuIlePro	312	
Db	102172	CGCACACCTTGCTGATTGGCCAGCGCAGCGGTGGAAATCGTCGAGCTAGAGCCC	102231	
Qy	313	AlaGlyGlyAlaAlaIleThr-----GluPheLysValAspValPro	326	
Db	102232	-----GGCGAGACGTCACCTTGAATCTGATGTTTGAAGACAATAACGCGCTCCCT	102285	
Qy	327	GlyAspTYrValLeuValAspHisAlaIlePheArgAlaPheAsnLysGlyAlaLeuGly	346	
Db	102286	GATGATGAGTCTGTGCTCCGATTTCCGACATGCTAGATTCCTTCCAG-----	102330	
Qy	347	IleLeuLysValGluGlyGluGluAsnHisGluIleTYrSerHisLysGlnThrAspAla	366	
Db	102331	CTGCTCACCATCATCCGCGCTTCCGAT-----GATGCT	102363	
Qy	367	ValTYrLeuPro-----GluGlyAlaProGlnAlaIleAsp	378	
Db	102364	GCGCAGACACCTCTTTCGCGGCGCTGCTGTAATTCACCGAACCTGACGTCACTCGAT	102423	
Qy	379	ThrGlnGluAlaProLys-----ThrProAlaProAlaLeuGlnGluGlnIleLys	396	
Db	102424	GCCACTGAACGCACTTCACTATCAATGAACACCTTCTCATCAACGATC-----	102469	
Qy	396	SalAGlyLysAlaThrTYrAspSerAsnCysAlaAla	408	
Db	102470	-----TACAGATGAGCATCGACGCGG	102490	
RESULT 39				
ID	AAT63317	standard; DNA; 1588 BP.		
XX	AAT63317;			
XX	07-AUG-1997	(first entry)		
XX	Rhizoctonia solani	laccase isozyme 4 (rs14) gene.		
KM	Blue copper oxidase; laccase; enzyme engineering;			
KW	protein engineering; lignin depolymerisation; dye oxidation; ss.			
XX	Rhizoctinia solani.			
OS				
XX				
XX	Key	Location/Qualifiers		
XX	CDS	1135..1137		
FT		/tag= a		
FT		/transl_except= pos:1135..1137;-aa:Lys		

FT		/note= "ATC codes for Ile"
FT	CDS	1243..1245
FT		/*tag= b
FT		/transl_except= pos:1243..1245; aa:Ile
FT		/note= "AAA codes for Lys"
XX	PN	M09709431-A1.
XX	PD	13-MAR-1997.
XX	PF	03-SEP-1996; 96MO-US14087.
XX	PR	01-SEP-1995; 95US-0003142.
XX	PA	(NOVO ) NOVO NORDISK BIOTECH INC.
XX	PI	Berka RM, Wahlechner JA, Xu F, Berka R;
XX	DR	WP1: 1997-192906/17.
XX	DR	P-PSDB; AAWI6301.
PT		New mutant blue copper oxidase enzymes - having different specific
PT		activities to wild-type enzymes, used for e.g. liginin
PT		(de) polymerisation or oxidation of dyes
XX		
PS		Disclosure; Fig 6A-D; 48pp; English.
CC		The rml4 gene (AAW63317) encoding Rhizoctonia solani laccase isozyyme
CC		4 (AAW16301) can be subjected to site-directed mutagenesis in order to
CC		alter e.g. the specific activity or pH-activity profile of the
CC		enzyme, or to improve expression yields. The mutation is a
CC		deletion, insertion or pref. a substn. of one or more amino acids
CC		at a location no greater than 12 Angstroms, pref. no more than 2.5
CC		Angstroms, from the Type I copper site. The mutant enzyme is
CC		expressed in transformed host cells for use e.g. in the
CC		polymerisation or depolymerisation of lignin, oxidation of dyes,
CC		etc.
XX		
SQ		Sequence 1588 BP; 362 A; 478 C; 376 G; 372 T; 0 other;
Alignment Scores:		
Pred. No.:	0.00481	Length: 1588
Score:	141.00	Matches: 103
Percent Similarity:	32.28%	Conservative: 60
Best Local Similarity:	20.40%	Mismatches: 216
Query Match:	5.37%	Indels: 126
DB:	18	Gaps: 20
US-10-088-045-2 (1-502) x AAW63317 (1-1588)		
Dy	59	LcuProValIIleAspAlalIleValThrhIShlaProGIuValProProProValAspArg 78
Dd	22	CTACCTTGTGCTGGCGGTCAACC----- 48
Dy	79	AspHisProAlaLysValValValLys-----MetGIutrhValGIuLysValMetArg 96
Dd	49	-----CCGCSCTTTGTCGCCGCAACTATAAGTTGCAGATCAAGAAGCATGTC 102
Dy	97	IeuAlaAspGIyValGIuTrcInIethrPrhRheGIyGIuInValProGIyGIuMet 116
Dd	103	GCTCCGAATGGCTTTCAGGCGCTTAATGCTCCGTCGAACGGATTAAATTCGGACGGTg 162
Dy	117	IleArgValArgGIuGIyAspThrIleGIuValGIuPheSerAsnHis-----Pro 133
Dd	163	ATACGGCCACAACAGGCTGACACCTTGGCATTAATGACAGATCAACTCACGGACCT 222
Dy	134	AspSerLysMetProHisAsnValAspPheHis-----AlaIalThrGIyPro 149
Dd	223	AGTATGCTGTGCACCAACGATTCATTCAGCATGATGTTCCAAGCTACTACCGCGAC 282
Dy	150	GIyGIyGIyAlaGIuIalSerPheThrIalPro-----GIyHisThrSer 166
Dd	283	GAGGATGGCCCGCATTCGTACGCAATGCCCTATGCGCAAAATTTGCTTATCAATAC 342



US-10-088-045-2 (1-502) x AA224235 (1-1588)

QY 59 LeuprovalilLeuAspAlaIleValThrHisAlaProGluValProProProValAspArg 78  
 DB 22 CTACCTTTCCTGCTCGGCTCAAC----- 48  
 QY 79 AspHisProAlaIleValValIleValHis-----MetGluThrValGluIleValMetArg 96  
 DB 49 -----CCGCGCTTTCCTGCTCGGCTCAACCTATATAGTCGACATCAAGAACGTCATGTC 102  
 QY 97 LeuAlaAspGlyValGluIleValThrHisProGluValProGluValProGluValMet 116  
 DB 103 GCTCCGATGCTGCTTCAGCGCTTCATGCTTCGTCACGCTTATAGTTTCTGCGACGCTG 162  
 QY 117 IleArgValAlaGluIleValAspThrIleGluValGlnPheSerHis-----Pro 133  
 DB 163 ATCAGCGCCCAACAAAGGATGACCTTGCAGCTTATATGTCACGATCACTACGACGACCT 222  
 QY 134 AspSerLysMetProHisAsnValAspPheHis-----AlaAlaThrGlyPro 149  
 DB 222 AGTATGCTGCTGCGCCCAACAGATTCATGCGATGATGTTCCAACTACTACGCGGAC 282  
 QY 150 GlyGlyGlyAlaGluAlaSerPheThrAlaPro-----GlyHisThrSer 164  
 DB 283 GAGGATGCGCGCCGATTCGTCACGCAATGCCCTATGCGCAAAATTGCTCTATACATAC 342  
 QY 165 ThrPheSerPheLysAlaLeuGlnProGluIleValThrHisCysAlaValAlaPro 184  
 DB 343 GAGATCCCATGTCGCGCC-----CAACAGAAACCATGTGATGAC-----GCCCAT 390  
 QY 185 ValGlyMetHisIleAlaAsnGlyMetGly-----LeuIleLeuValGluPro----- 201  
 DB 391 CTTCGAGTCATATGTCATGATGATGTCGAGGCGCTTGGTCTCATATGATCCAAACGAC 450  
 QY 202 -----LysGluGlyLeuProLysValAspLysGlyValThrValMetGlnLysAspPhe 219  
 DB 451 CCACACAACTCCGCTACGACGATGATGATGACGAGCAGATGATGATGATGATGATGATG 510  
 QY 220 Tyr-----ThrLysGlyLysValGlyGluGluGlu----- 229  
 DB 511 TACCATCTCCGCGACCCGCTTCTAGAAAGCAAAATGTTCTCGCATATTAACACCGCTCTG 570  
 QY 230 LeuGlnProPheAspMetGluLysAlaIleArgGluAspAlaGluValValPheAsn 249  
 DB 571 CTCTCTCTGTCGCGACCTCGGCTCTTATCAATGCGCAAGGCGCTATG----- 621  
 QY 250 GlySerValGlyAlaLeuThrGlyGluAsnAlaLeuLysValGlyGluThrVal 269  
 DB 622 -----GCGCGTCCCGCAGTCCCGCTCAGTATCAACGTAAACGTGGAAACGATAT 675  
 QY 270 ArgLeuPheValGlyAsnGlyGlyProAsnLeuThrSerSerPheHisValIleGlyGlu 289  
 DB 676 CGCTTGGCGGTATCAACGCTTCTGATCGGCTCTTACCTTTTCGATC----- 726  
 QY 290 IlePheAspLysValHisPheGluGlyGlyLysGlyGluAsnHisAsnIleGlnThr 309  
 DB 727 -----GAAAGCATAGTCTG-----ACT 744  
 QY 310 LeuIleProAlaGlyGly-----AlaAlaIleThrGluPheLysValAsp 324  
 DB 745 GTCATTTGAGGCGGATGCTGTCACAGCCCTTGGCTGCTTACAGCTTCCAGATTTAC 804  
 QY 325 ValProGlyAspArgValLeuValAspHisAlaIlePheArgAlaPheAsnLysGlyAla 344  
 DB 805 GCTGCAACAGCTACTCTGTCAGCTTGAAGCCAAACCAACCGCGCCCAACTACTGAT 864  
 QY 345 LeuGlyIleLeuValGluGlyGluGluAsnHisGluIleTyrSerHisLysGlnThr 364  
 DB 865 CGTGACCAATGACCTTGGAGAGCGCGCAACCAATGCAACTTGGACCCCACTATGTC 924  
 QY 365 AspAlaValIleValProGluGlyAlaProGlnAlaIleAspThrGlnGlu----- 381

DB 925 TTTCGCGTATTCGATCAGGAGGAGCGCCCAAGCCGAAACCAACGACGACAGGACCT 984  
 QY 382 -----AlaPro 383  
 DB 985 GGTATGCGTACTGCACTGCTTGAAGAACTCCATGCGCTGATCAACCTGCGCTCG 1044  
 QY 384 LysThrProAlaProAlaAsnLeuGlnGluGlnIleLysAlaGlyLysAlaThrTyrAsp 403  
 DB 1045 GCGGCTCCGCTCCGCAACGCTTCCCTCAATCTTGCAATTGGCGGACAGCTTGAT 1104  
 QY 404 Ser-----AsnCysAlaIleCysHisGlnProAspLysGlyVal 417  
 DB 1105 GGGATTTTGAAGTTCATTAATTAATCAATCAATCAAGAGCTCT----- 1149  
 QY 418 ProAsnAlaPheProProLeuAlaAsnSerAspTyrLeuAsnAlaAspHisAlaArgAla 437  
 DB 1150 -----TCGTTGCCACGCTC----- 1164  
 QY 438 AlaSerIleValAlaAsnGlyLeuSerGlyLysIleThrValAsnGlyAsnGlnTyrGlu 457  
 DB 1165 TTGAAGATTTTGGCAACATGCGAGCAATGACGCGGATTTCAAGCCCAATGACACACT 1224  
 QY 458 SerValMetProAlaIleAlaLeuSerAspGlnGlnIleAlaAsnValIleThrTyr 477  
 DB 1225 ATCGATTTGCGCACAAATTAAGTTATCGAGCTCAATATACCGGAGTGCAGACACCT 1284  
 QY 478 LeuAsnSerPheGlyAsnLysGlyGlyGlnLeuSerAlaAspValAlaLysAlaLys 497  
 DB 1285 ATCCATCTCCAGGCGCAT-----GTGTTGATATCGTCAAATCAGTC 1326  
 QY 498 LysThrLysProAsn 502  
 DB 1327 GGTGTATCCCGAAC 1341

Search completed: September 5, 2003, 14:35:31  
 Job time : 13762 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 27, 2003, 18:40:47 ; Search time 29 Seconds

(without alignments)  
732.416 Million cell updates/sec

Title: US-10-088-045-2

Perfect score: 502

Sequence: 1 MSKPTLTKTICALSALML.....NKGQLSDADYAKKTKRN 502

Scoring table: OLIGO

Searched: 328717 seqs, 42310858 residues

Word size: 0

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database: Issued Patents\_AA:\*

1: /cgn2\_6/pdata/1/1aa/5A\_COMB.pep:\*  
2: /cgn2\_6/pdata/1/1aa/5B\_COMB.pep:\*  
3: /cgn2\_6/pdata/1/1aa/6A\_COMB.pep:\*  
4: /cgn2\_6/pdata/1/1aa/6B\_COMB.pep:\*  
5: /cgn2\_6/pdata/1/1aa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/pdata/1/1aa/Backfill.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7	1.4	33	US-09-433-428D-40	Sequence 40, Appl
2	7	1.4	104	US-08-817-441-100	Sequence 100, App
3	7	1.4	172	US-09-134-001C-2971	Sequence 2971, Ap
4	7	1.4	177	US-09-252-991A-18782	Sequence 18782, A
5	7	1.4	180	US-09-173-300-28	Sequence 28, Appl
6	7	1.4	180	US-09-252-991A-22780	Sequence 22780, A
7	7	1.4	215	US-08-311-731A-410	Sequence 410, App
8	7	1.4	224	US-09-082-920-2	Sequence 2, Appl
9	7	1.4	233	US-09-214-631-7	Sequence 7, Appl
10	7	1.4	234	US-08-299-567-5	Sequence 5, Appl
11	7	1.4	238	US-08-240-124-2	Sequence 2, Appl
12	7	1.4	238	US-08-453-943-2	Sequence 2, Appl
13	7	1.4	238	US-09-057-121-2	Sequence 2, Appl
14	7	1.4	238	US-09-358-734-2	Sequence 2, Appl
15	7	1.4	262	US-09-252-991A-22061	Sequence 22061, A
16	7	1.4	300	US-09-252-991A-27792	Sequence 27792, A
17	7	1.4	312	US-08-821-872-2	Sequence 2, Appl
18	7	1.4	312	US-09-171-517B-11	Sequence 11, Appl
19	7	1.4	320	US-09-134-001C-4439	Sequence 4439, Ap
20	7	1.4	326	US-09-154-874-9	Sequence 9, Appl
21	7	1.4	326	US-08-931-668-9	Sequence 9, Appl
22	7	1.4	327	US-09-154-874-8	Sequence 8, Appl
23	7	1.4	327	US-08-931-668-8	Sequence 8, Appl
24	7	1.4	376	US-09-252-991A-21843	Sequence 21843, A
25	7	1.4	438	US-08-886-886-2	Sequence 2, Appl
26	7	1.4	438	US-09-044-718-2	Sequence 26, Appl
27	7	1.4	442	US-09-296-284-26	Sequence 26, Appl

28	7	1.4	450	US-09-044-718-15	Sequence 15, Appl
29	7	1.4	466	US-08-868-433-35	Sequence 35, Appl
30	7	1.4	466	US-08-744-231-35	Sequence 35, Appl
31	7	1.4	478	US-09-296-284-5	Sequence 5, Appl
32	7	1.4	518	US-08-836-620A-18	Sequence 18, Appl
33	7	1.4	521	US-08-878-563A-3	Sequence 3, Appl
34	7	1.4	521	US-09-270-117-3	Sequence 3, Appl
35	7	1.4	534	US-08-878-563A-1	Sequence 1, Appl
36	7	1.4	539	US-09-270-117-1	Sequence 1, Appl
37	7	1.4	539	US-09-328-352-5771	Sequence 5771, Ap
38	7	1.4	580	US-09-252-991A-20407	Sequence 20407, A
39	7	1.4	616	US-09-252-991A-19164	Sequence 19164, A
40	7	1.4	727	US-09-252-991A-16100	Sequence 26100, A
41	7	1.4	962	US-09-614-480-9	Sequence 9, Appl
42	7	1.4	988	US-09-614-480-2	Sequence 2, Appl
43	7	1.4	1190	US-09-107-532A-7146	Sequence 7146, Ap
44	7	1.2	9	US-09-005-215-13	Sequence 13, Appl
45	6	1.2	9	US-09-344-040C-117	Sequence 117, Appl
46	6	1.2	10	US-08-453-924-16	Sequence 16, Appl
47	6	1.2	11	US-09-133-062D-28	Sequence 28, Appl
48	6	1.2	11	US-09-133-062D-29	Sequence 29, Appl
49	6	1.2	12	US-08-477-674-1	Sequence 1, Appl
50	6	1.2	12	US-08-473-791-1	Sequence 1, Appl
51	6	1.2	12	US-08-316-714-1	Sequence 1, Appl
52	6	1.2	12	US-08-473-791-11	Sequence 11, Appl
53	6	1.2	13	US-08-305-871A-18	Sequence 18, Appl
54	6	1.2	13	US-08-411-859-15	Sequence 15, Appl
55	6	1.2	13	US-08-788-822A-22	Sequence 22, Appl
56	6	1.2	13	US-09-068-150-1	Sequence 1, Appl
57	6	1.2	13	US-08-477-674-11	Sequence 11, Appl
58	6	1.2	23	US-08-473-791-11	Sequence 11, Appl
59	6	1.2	23	US-08-316-714-11	Sequence 11, Appl
60	6	1.2	23	US-08-473-673-11	Sequence 11, Appl
61	6	1.2	23	US-09-425-638A-82	Sequence 82, Appl
62	6	1.2	23	US-09-425-638A-83	Sequence 83, Appl
63	6	1.2	23	US-09-425-638A-84	Sequence 84, Appl
64	6	1.2	23	US-09-543-004-82	Sequence 82, Appl
65	6	1.2	23	US-09-543-004-83	Sequence 83, Appl
66	6	1.2	23	US-09-543-004-84	Sequence 84, Appl
67	6	1.2	31	US-09-071-224-36	Sequence 36, Appl
68	6	1.2	37	US-09-082-278B-1216	Sequence 1216, Ap
69	6	1.2	37	US-09-315-304B-1216	Sequence 1216, Ap
70	6	1.2	37	US-09-834-784-1216	Sequence 1216, Ap
71	6	1.2	41	US-08-453-924-7	Sequence 7, Appl
72	6	1.2	41	US-08-129-129-5	Sequence 5, Appl
73	6	1.2	46	US-08-415-751-13	Sequence 13, Appl
74	6	1.2	47	US-08-469-268A-556	Sequence 556, App
75	6	1.2	47	US-08-488-446-556	Sequence 556, App
76	6	1.2	47	US-08-467-344A-556	Sequence 556, App
77	6	1.2	53	US-08-858-207A-444	Sequence 444, App
78	6	1.2	55	US-08-468-580-11	Sequence 11, Appl
79	6	1.2	55	US-08-468-580-12	Sequence 12, Appl
80	6	1.2	55	US-08-468-580-15	Sequence 15, Appl
81	6	1.2	55	US-08-468-580-16	Sequence 16, Appl
82	6	1.2	55	US-08-643-212-10	Sequence 10, Appl
83	6	1.2	55	US-08-643-212-12	Sequence 12, Appl
84	6	1.2	55	US-08-643-212-14	Sequence 14, Appl
85	6	1.2	55	US-08-643-212-16	Sequence 16, Appl
86	6	1.2	55	PCT-US95-03731-11	Sequence 11, Appl
87	6	1.2	55	PCT-US95-03731-12	Sequence 12, Appl
88	6	1.2	55	PCT-US95-03731-15	Sequence 15, Appl
89	6	1.2	55	PCT-US95-03731-16	Sequence 16, Appl
90	6	1.2	72	US-09-328-352-566A	Sequence 6264, Ap
91	6	1.2	74	US-09-252-991A-22657	Sequence 22657, A
92	6	1.2	74	PCT-US92-00282-24	Sequence 24, Appl
93	6	1.2	75	US-08-604-989A-2	Sequence 2, Appl
94	6	1.2	77	PCT-US95-02795A-4	Sequence 4, Appl
95	6	1.2	79	US-08-904-263A-6	Sequence 6, Appl
96	6	1.2	79	US-09-434-123A-6	Sequence 6, Appl
97	6	1.2	79	PCT-US95-06266-132	Sequence 132, App
98	6	1.2	80	PCT-US95-06266-114	Sequence 114, App
99	6	1.2	80	PCT-US95-06266-118	Sequence 118, App
100	6	1.2	80	PCT-US95-06266-120	Sequence 120, App















977 5 1.0 10 1 US-08-003-966-1 Sequence 1, Appl  
978 5 1.0 10 1 US-07-940-861-3 Sequence 3, Appl  
979 5 1.0 10 1 US-08-214-650-16 Sequence 16, Appl  
980 5 1.0 10 1 US-08-459-512-3 Sequence 3, Appl  
981 5 1.0 10 1 US-08-627-497-7 Sequence 7, Appl  
982 5 1.0 10 2 US-08-572-951-34 Sequence 34, Appl  
983 5 1.0 10 2 US-08-533-298-17 Sequence 17, Appl  
984 5 1.0 10 2 US-08-459-657-3 Sequence 3, Appl  
985 5 1.0 10 2 US-08-460-132-3 Sequence 3, Appl  
986 5 1.0 10 3 US-08-159-339A-145 Sequence 145, App  
987 5 1.0 10 3 US-08-602-999A-5 Sequence 5, Appl  
988 5 1.0 10 4 US-08-278-865-5 Sequence 5, Appl  
989 5 1.0 10 4 US-09-500-124-5 Sequence 5, Appl  
990 5 1.0 10 4 US-09-106-872A-11 Sequence 11, Appl  
991 5 1.0 10 5 PCT-US92-02050-3 Sequence 3, Appl  
992 5 1.0 11 3 US-08-602-999A-262 Sequence 262, App  
993 5 1.0 11 3 US-08-602-999A-290 Sequence 290, App  
994 5 1.0 11 3 US-08-652-877-24 Sequence 24, App  
995 5 1.0 11 3 US-08-476-515A-24 Sequence 24, App  
996 5 1.0 11 4 US-09-500-124-262 Sequence 262, App  
997 5 1.0 11 4 US-09-500-124-290 Sequence 290, App  
998 5 1.0 11 6 5187155-26 Patent No. 5187155  
999 5 1.0 12 2 US-08-659-567-8 Patent No. 5187155  
1000 5 1.0 12 2 US-08-538-960-4 Sequence 4, Appl

## ALIGNMENTS

RESULT 1  
US-09-433-428D-40  
; Sequence 40, Application US/09433428D  
; Patent No. 6149910  
; GENERAL INFORMATION:  
; APPLICANT: De Leys, Robert J.  
; APPLICANT: Zheng, Jian  
; TITLE OF INVENTION: PEPTIDES FOR THE DETECTION OF HIV-1 GROUP O  
; FILE REFERENCE: CDS-207  
; CURRENT APPLICATION NUMBER: US/09/433,428D  
; CURRENT FILING DATE: 1999-11-04  
; NUMBER OF SEQ ID NOS: 70  
; SOFTWARE: Patentn Ver. 2.0  
; SEQ ID NO 40  
; LENGTH: 33  
; TYPE: PRT  
; ORGANISM: Human immunodeficiency virus type 1  
US-09-433-428D-40

Query Match 1.4%; Score 7; DB 3; Length 33;  
Best Local Similarity 100.0%; Pred. No. 16;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 467 DOQIANV 473  
Db 1 DOQIANV 7

RESULT 2  
US-08-817-441-100  
; Sequence 100, Application US/08817441  
; Patent No. 6399294  
; GENERAL INFORMATION:  
; APPLICANT: CHARNEAU, PIERRE  
; APPLICANT: CLAVEL, FRANCOISE  
; APPLICANT: BORMAN, ANDREW  
; APPLICANT: QUILLENT, CAROLINE  
; APPLICANT: GUETARD, DENISE  
; APPLICANT: MONTAGNIER, LUC  
; APPLICANT: DONJON DE SAINT-MARTIN, JACQUELINE  
; APPLICANT: COHEN, JACQUES  
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES OF HIV-1 TYPE (OR  
; TITLE OF INVENTION: SUBTYPE) ANTIGENS  
; NUMBER OF SEQUENCES: 103

CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &  
; ADDRESSEE: Dunner, L.L.P.  
; STREET: 1300 I Street, N.W.  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20005-3315  
COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/817,441  
; FILING DATE: 11-JUL-1997  
PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/FR 95/01391  
; FILING DATE: 20-OCT-1995  
PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: FR 9412554  
; FILING DATE: 20-OCT-1994  
PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: FR 9502526  
; FILING DATE: 03-MAR-1995  
ATTORNEY/AGENT INFORMATION:  
; NAME: Meyers, Kenneth J.  
; REGISTRATION NUMBER: 25,146  
REFERENCE/DOCKET NUMBER: 03260, 6005-00000  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-408-4000  
; TELEFAX: 202-408-4400  
INFORMATION FOR SEQ ID NO: 100:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 104 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-817-441-100

Query Match 1.4%; Score 7; DB 4; Length 104;  
Best Local Similarity 100.0%; Pred. No. 43;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 467 DOQIANV 473  
Db 83 DOQIANV 89

RESULT 3  
US-09-134-001C-2971  
; Sequence 2971, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 2971  
; LENGTH: 172  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-2971

Query Match 1.4%; Score 7; DB 4; Length 172;

Best Local Similarity 100.0%; Pred. No. 67;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 219 FTTKGY 225  
DB 151 FTTKGY 157

## RESULT 4

US-09-252-991A-18782  
; Sequence 18782, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 18782  
; LENGTH: 177  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-18782

Query Match  
Best Local Similarity 100.0%; Pred. No. 68;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 146 ATGPGGG 152  
DB 171 ATGPGGG 177

## RESULT 5

US-09-173-300-28  
; Sequence 28, Application US/09173300  
; Patent No. 6451581  
; GENERAL INFORMATION:  
; APPLICANT: Falco, Saverio Carl  
; APPLICANT: Hiltz, William D.  
; APPLICANT: Kinney, Anthony J.  
; APPLICANT: Cahoon, Rebecca E.  
; APPLICANT: Rafalski, J. Antoni  
; TITLE OF INVENTION: PLANT BRANCHED CHAIN AMINO ACID BIOSYNTHETIC ENZYMES  
; FILE REFERENCE: BB-1126  
; CURRENT APPLICATION NUMBER: US/09/173,300  
; CURRENT FILING DATE: 1998-10-15  
; EARLIER APPLICATION NUMBER: 60/063,423  
; EARLIER FILING DATE: 1997 October 28  
; NUMBER OF SEQ ID NOS: 54  
; SOFTWARE: Microsoft Word Version 7.0A  
; SEQ ID NO 28  
; LENGTH: 180  
; TYPE: PRT  
; ORGANISM: Triticum aestivum  
US-09-173-300-28

Query Match  
Best Local Similarity 100.0%; Pred. No. 69;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 349 KVEGEEN 355  
DB 58 KVEGEEN 64

## RESULT 6

US-09-252-991A-22780  
; Sequence 22780, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 22780  
; LENGTH: 180  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-22780

Query Match  
Best Local Similarity 100.0%; Pred. No. 69;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MSKPTLI 7  
DB 119 MSKPTLI 125

RESULT 7  
US-08-311-731A-410  
; Sequence 410, Application US/08311731A  
; Patent No. 6583266  
; GENERAL INFORMATION:  
; APPLICANT: SMITH, DOUGLAS  
; APPLICANT: MAO, JEN-I  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES  
; TITLE OF INVENTION: RELATING TO MYCOBACTERIUM TUBERCULOSIS AND LAPRAE FOR  
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS  
; NUMBER OF SEQUENCES: 411  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.  
; STREET: 600 ATLANTIC AVENUE  
; CITY: BOSTON  
; STATE: MASSACHUSETTS  
; COUNTRY: USA  
; ZIP: 02210

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/311,731A  
FILING DATE:  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: GATES, EDWARD R.  
REGISTRATION NUMBER: 31,616  
REFERENCE/DOCKET NUMBER: C0044/7125  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617/720-3500  
TELEFAX: 617/720-2441  
INFORMATION FOR SEQ ID NO: 410:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 215 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ORIGINAL SOURCE:  
ORGANISM: Mycobacterium leprae

US-08-311-731A-410

Query Match	1.4%	Score 7;	DB 4;	Length 215;
Best Local Similarity	100.0%	Pred. No. 81;		
Matches	7;	Conservative	0;	Mismatches 0;
				Indels 0;
				Gaps 0;

QY	61	VIDAIVT	67
Db	104	VIDAIVT	110

RESULT 8  
US-09-082-920-2  
; Sequence 2, Application US/09082920C  
; Data No. 5355460

```

; APPLICANT: Lam, Kelvin T.
; TITLE OF INVENTION: Nucleic Acid Encoding M. Tuberculosis ALGU Protein
; FILE REFERENCE: 0342/1C587USAPPEND
; CURRENT APPLICATION NUMBER: US/09/082,920C
; CURRENT FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 60/035,391
; EARLIER FILING DATE: 1997-01-16
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO: 2
; LENGTH: 224
; TYPE: PRT
; ORGANISM: Mycobacteria tuberculosis
US-09-082-920-2

```

Query Match	1.4%	Score 7;	DB 4;	Length 224;
Best Local Similarity	100.0%	Pred. No. 84;		
Matches	7;	Conservative	0;	Mismatches 0;
			Indels	0;
			Gaps	0;

QY	488	LSADDVA	49
Db	83	LSADDVA	89

```

RESULT 9
US-09-214-631-7
; Sequence 7, Application US/09214631
; Patent No. 6413730
; GENERAL INFORMATION:
; APPLICANT: Holland, Sacha
; APPLICANT: Mbamalu, Geraldine
; APPLICANT: Pearson, Tony
; TITLE OF INVENTION: OLIGOMERIZED RECEPTORS WHICH AFFECT PATHWAYS REGULATED
; TITLE OF INVENTION: BY TRANSMEMBRANE LIGANDS FOR ELK-RELATED RECEPTOR
; TITLE OF INVENTION: TYROSINE KINASES
; FILE REFERENCE: 11757.23USMO
; CURRENT APPLICATION NUMBER: US/09/214,631
; CURRENT FILING DATE: 1999-03-12
; EARLIER APPLICATION NUMBER: PCT/CA97/00473
; EARLIER FILING DATE: 1997-07-04
; EARLIER APPLICATION NUMBER: 60/021,272
; EARLIER FILING DATE: 1996-07-05
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 233
; TYPE: prt
; ORGANISM: Homo sapiens
US-09-214-631-7

```

Query Match	1.4%;	Score 7;	DB 4;	Length 233;
Best Local Similarity	100.0%;	Pred. No. 87;		
Matches	7;	Conservative	0;	Mismatches 0;
				Indels 0;
				Gaps 0;

```

QY      148 GPGGGAE 15
          |||||
Db      74  GPGGGAE 80

```

RESULT 10  
US-08-299-567-5  
Continued on Page 10/0020000000

APPLICANT: Davis, et al.  
TITLE OF INVENTION: METHOD OF ENHANCING THE BIOLOGICAL  
TITLE OF INVENTION: ACTIVITY OF EPH FAMILY LIGANDS  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Regeneron Pharmaceuticals, Inc.  
STREET: 777 Old Saw Mill River Road  
CITY: Tarrytown  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10591-6707  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/299,567  
FILING DATE: 01-SEP-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Kempfner, Gail M.  
REGISTRATION NUMBER: 32,143  
REFERENCE/DOCKET NUMBER: REG 290  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 914-345-7400  
TELEFAX: 914-345-7721  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 234 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-299-567-5

Query Match	1.4%	Score 7	DB 1	Length 234
Similarity	100.0%	Pred. No. 87		
Best Local	7	Conservative	0	Indels 0
Matches				Gaps 0

QY	148	GGGGAE	154
Db	74	GGGGAE	80

RESULT 11-11-21  
US-08-240-124-2  
Sequence 2, Application US/08240124  
Patent No. 511658  
GENERAL INFORMATION:  
APPLICANT: BECKMANN, M. P.  
APPLICANT: CERRETTI, DOUGLAS P.  
TITLE OF INVENTION: CYTOKINE THAT BINDS THE CELL SURFACE  
TITLE OF INVENTION: RECEPTOR HEM  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: IMMUNEX CORPORATION  
STREET: 51 UNIVERSITY STREET  
CITY: SEATTLE  
STATE: WASHINGTON  
COUNTRY: USA  
ZIP: 98101  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: Apple Macintosh  
OPERATING SYSTEM: Apple System 7.1



SOFTWARE: Microsoft Word for Apple, Version 5.1a  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/240,124  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/161,132  
FILING DATE: 03-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/114,426  
FILING DATE: 30-AUG-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/109,745  
FILING DATE: 20-AUG-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: SEESE, KATHRYN A.  
REGISTRATION NUMBER: 32,172  
REFERENCE/DOCKET NUMBER: 2814-C  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 587-0430  
TELEFAX: (206) 233-0644  
TELEX: 756822  
INFORMATION FOR SEQ. ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 238 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-240-124-2

Query Match 1.4%; Score 7; DB 1; Length 238;  
Best Local Similarity 100.0%; Pred. No. 89;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 148 GPGGAE 154  
|||||  
Db 78 GPGGAE 84

RESULT 12  
US-08-453-943-2  
Sequence 2, Application US/08453943  
Patent No. 5738844  
GENERAL INFORMATION:  
APPLICANT: BECKMANN, M. P.  
TITLE OF INVENTION: CYTOKINE THAT BINDS THE CELL SURFACE  
TITLE OF INVENTION: RECEPTOR HEK  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: IMMUNEX CORPORATION  
STREET: 51 UNIVERSITY STREET  
CITY: SEATTLE  
STATE: WASHINGTON  
COUNTRY: USA  
ZIP: 98101  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: Apple Macintosh  
OPERATING SYSTEM: Apple System 7.1  
SOFTWARE: Microsoft Word for Apple, Version 5.1a  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/453,943  
FILING DATE: 30-MAY-1995  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/240,124  
FILING DATE: 09-MAY-1994  
APPLICATION NUMBER: US 08/161,132  
FILING DATE: 03-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/114,426  
FILING DATE: 30-AUG-1993

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/109,745  
FILING DATE: 20-AUG-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: SEESE, KATHRYN A.  
REGISTRATION NUMBER: 32,172  
REFERENCE/DOCKET NUMBER: 2814-C  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 587-0430  
TELEFAX: (206) 233-0644  
TELEX: 756822  
INFORMATION FOR SEQ. ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 238 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-453-943-2

Query Match 1.4%; Score 7; DB 1; Length 238;  
Best Local Similarity 100.0%; Pred. No. 89;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 148 GPGGAE 154  
|||||  
Db 78 GPGGAE 84

RESULT 13  
US-09-057-121-2  
Sequence 2, Application US/09057121  
Patent No. 5969110  
GENERAL INFORMATION:  
APPLICANT: BECKMANN, M. P.  
TITLE OF INVENTION: CYTOKINE THAT BINDS THE CELL SURFACE  
TITLE OF INVENTION: RECEPTOR HEK  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: IMMUNEX CORPORATION  
STREET: 51 UNIVERSITY STREET  
CITY: SEATTLE  
STATE: WASHINGTON  
COUNTRY: USA  
ZIP: 98101  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: Apple Macintosh  
OPERATING SYSTEM: Apple System 7.1  
SOFTWARE: Microsoft Word for Apple, Version 5.1a  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/057,121  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/240,124  
FILING DATE:  
APPLICATION NUMBER: US 08/161,132  
FILING DATE: 03-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/114,426  
FILING DATE: 30-AUG-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/109,745  
FILING DATE: 20-AUG-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: SEESE, KATHRYN A.  
REGISTRATION NUMBER: 32,172  
REFERENCE/DOCKET NUMBER: 2814-C  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 587-0430  
TELEFAX: (206) 233-0644  
TELEX: 756822

INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 238 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-057-121-2

Query Match 1.4%; Score 7; DB 2; Length 238;  
Best Local Similarity 100.0%; Pred. No. 89;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 148 GPGGAE 154  
DB 78 GPGGAE 84

RESULT 14  
US-09-358-734-2  
Sequence 2, Application US/09358734  
Patent No. 6274117  
GENERAL INFORMATION:  
APPLICANT: BECKMANN, M. P.  
TITLE OF INVENTION: CYTOKINE THAT BINDS THE CELL SURFACE  
TITLE OF INVENTION: RECEPTOR HEK  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESS: IMMUNEX CORPORATION  
STREET: 51 UNIVERSITY STREET  
CITY: SEATTLE  
STATE: WASHINGTON  
COUNTRY: USA  
ZIP: 98101  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: Apple Macintosh  
OPERATING SYSTEM: Apple System 7.1  
SOFTWARE: Microsoft Word for Apple, Version 5.1a  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/358,734  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/240,124  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/114,426  
FILING DATE: 30-AUG-1993  
APPLICATION NUMBER: US 08/109,745  
FILING DATE: 20-AUG-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: SEESE, KATHRYN A.  
REGISTRATION NUMBER: 32,172  
REFERENCE/DOCKET NUMBER: 2814-C  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 587-0430  
TELEFAX: (206) 233-0644  
TELEX: 756822  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 238 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-358-734-2

Query Match 1.4%; Score 7; DB 3; Length 238;  
Best Local Similarity 100.0%; Pred. No. 89;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 148 GPGGAE 154

DB 78 GPGGAE 84

RESULT 15  
US-09-252-991A-22061  
Sequence 22061, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 22061  
LENGTH: 262  
TYPE: PRT  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-22061

Query Match 1.4%; Score 7; DB 4; Length 262;  
Best Local Similarity 100.0%; Pred. No. 96;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 LMSGCS 24  
DB 213 LMSGCS 219

RESULT 16  
US-09-252-991A-27792  
Sequence 27792, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 27792  
LENGTH: 300  
TYPE: PRT  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-27792

Query Match 1.4%; Score 7; DB 4; Length 300;  
Best Local Similarity 100.0%; Pred. No. 116+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 435 ARAASIV 441  
DB 195 ARAASIV 201

RESULT 17  
US-08-821-872-2  
Sequence 2, Application US/08821872A  
Patent No. 6204004  
GENERAL INFORMATION:  
APPLICANT: Kaper, James B  
APPLICANT: Jarvis, Karen

;; TITLE OF INVENTION: Immunodiagnostic Test for Enterohemorrhagic Escherichia  
;; FILE OF INVENTION: coli Infection  
;; FILE REFERENCE: Seq. ID No. 62040044 Ref: 1419K  
;; CURRENT APPLICATION NUMBER: US/08/821,872A  
;; CURRENT FILING DATE: 1997-03-21  
;; NUMBER OF SEQ ID NOS: 2  
;; SOFTWARE: Patent In Ver. 2.1  
;; SEQ ID NO 2  
;; LENGTH: 312  
;; TYPE: PRT  
;; ORGANISM: Escherichia coli  
US-08-821-872-2

Query Match  
Best Local Similarity 1.4%; Score 7; DB 3; Length 312;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 490 ADDVAKA 496  
Db 187 ADDVAKA 193

RESULT 18  
US-09-171-517B-11  
;; Sequence 11, Application US/09171517B  
;; Patent No. 6355254  
;; GENERAL INFORMATION:  
;; APPLICANT: Finlay, B. Brett  
;; APPLICANT: Kenny, Brendan  
;; APPLICANT: Stein, Markus  
;; TITLE OF INVENTION: PATHOGENIC ESCHERICHIA COLI ASSOCIATED PROTEIN  
;; FILE REFERENCE: 07422/019001  
;; CURRENT APPLICATION NUMBER: US/09/171,517B  
;; CURRENT FILING DATE: 1998-10-20  
;; PRIOR APPLICATION NUMBER: PCT/CA97/00265  
;; PRIOR FILING DATE: 1997-04-23  
;; PRIOR APPLICATION NUMBER: 60/015,999  
;; PRIOR FILING DATE: 1996-04-23  
;; NUMBER OF SEQ ID NOS: 15  
;; SOFTWARE: Patent In Ver. 2.0  
;; SEQ ID NO 11  
;; LENGTH: 312  
;; TYPE: PRT  
;; ORGANISM: Escherichia coli  
US-09-171-517B-11

Query Match  
Best Local Similarity 1.4%; Score 7; DB 4; Length 312;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 490 ADDVAKA 496  
Db 187 ADDVAKA 193

RESULT 19  
US-09-134-001C-4439  
;; Sequence 4439, Application US/09134001C  
;; Patent No. 6380370  
;; GENERAL INFORMATION:  
;; APPLICANT: Lynn Doucette-Stamm et al  
;; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
;; FILE REFERENCE: GTC-007  
;; CURRENT APPLICATION NUMBER: US/09/134,001C  
;; CURRENT FILING DATE: 1998-08-13  
;; PRIOR APPLICATION NUMBER: US 60/064,964  
;; PRIOR FILING DATE: 1997-11-08  
;; PRIOR APPLICATION NUMBER: US 60/055,779  
;; PRIOR FILING DATE: 1997-08-14  
;; NUMBER OF SEQ ID NOS: 5674  
;; SEQ ID NO 4439  
;; LENGTH: 320

;; TYPE: PRT  
;; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-4439

Query Match  
Best Local Similarity 1.4%; Score 7; DB 4; Length 320;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 LMLSGCS 24  
Db 17 LMLSGCS 23

RESULT 20  
US-09-154-874-9  
;; Sequence 9, Application US/09154874  
;; Patent No. 6054636  
;; GENERAL INFORMATION:  
;; APPLICANT: FADER, GARY MICHAEL  
;; TITLE OF INVENTION: ISOFLAVONE BIOSYNTHETIC ENZYMES  
;; NUMBER OF SEQUENCES: 9  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY  
;; STREET: 1007 MARKET STREET  
;; CITY: WILMINGTON  
;; STATE: DELAWARE  
;; COUNTRY: UNITED STATES OF AMERICA  
;; ZIP: 19898  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: DISKETTE, 3.50 INCH  
;; COMPUTER: IBM PC COMPATIBLE  
;; OPERATING SYSTEM: MICROSOFT WINDOWS 95  
;; SOFTWARE: MICROSOFT WORD VERSION 7.0A  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/09/154,874  
;; FILING DATE:  
;; CLASSIFICATION:  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 08/931,668  
;; FILING DATE: SEPTEMBER 17, 1997  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: MAJARIAN, WILLIAM R.  
;; REGISTRATION NUMBER: 41,173  
;; REFERENCE/DOCKET NUMBER: BB-1098-A  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 302-992-4926  
;; TELEFAX: 302-773-0164  
;; INFORMATION FOR SEQ ID NO: 9:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 326 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS:  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
US-09-154-874-9

Query Match  
Best Local Similarity 1.4%; Score 7; DB 3; Length 326;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 343 GALGLIK 349  
Db 107 GALGLIK 113

RESULT 21  
US-08-931-668-9  
;; Sequence 9, Application US/08931668  
;; Patent No. 6521433  
;; GENERAL INFORMATION:  
;; APPLICANT: FADER, GARY M.  
;; TITLE OF INVENTION: CDNA SEQUENCES FROM SOYBEAN THAT ENCODE  
;; ACTIVITIES ASSOCIATED WITH ISOFLAVONE

TITLE OF INVENTION: BIOSYNTHESIS  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY  
STREET: 1007 MARKET STREET  
CITY: WILMINGTON  
STATE: DELAWARE  
COUNTRY: UNITED STATES OF AMERICA  
ZIP: 19898  
COMPUTER READABLE FORM:  
MEDIUM TYPE: DISKETTE, 3.50 INCH  
COMPUTER: IBM PC COMPATIBLE  
OPERATING SYSTEM: MICROSOFT WORD FOR WINDOWS 95  
SOFTWARE: MICROSOFT WORD VERSION 7.0A  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/931,668  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: MAJARIAN, WILLIAM R.  
REGISTRATION NUMBER: 41,173  
REFERENCE/DOCKET NUMBER: BB-1098  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 302-992-4926  
TELEFAX: 302-773-0164  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 326 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-931-668-9

Query Match 1.4%; Score 7; DB 4; Length 326;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 343 GAGILK 349  
DB 107 GAGILK 113

RESULT 22  
US-09-154-874-8  
Sequence 8, Application US/09154874  
Patent No. 6034636  
GENERAL INFORMATION:  
APPLICANT: FADER, GARY MICHAEL  
TITLE OF INVENTION: ISOPOLYMER BIOSYNTHETIC ENZYMES  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY  
STREET: 1007 MARKET STREET  
CITY: WILMINGTON  
STATE: DELAWARE  
COUNTRY: UNITED STATES OF AMERICA  
ZIP: 19898  
COMPUTER READABLE FORM:  
MEDIUM TYPE: DISKETTE, 3.50 INCH  
COMPUTER: IBM PC COMPATIBLE  
OPERATING SYSTEM: MICROSOFT WORDS 95  
SOFTWARE: MICROSOFT WORD VERSION 7.0A  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/154,874  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/931,668  
FILING DATE: SEPTEMBER 17, 1997  
ATTORNEY/AGENT INFORMATION:  
NAME: MAJARIAN, WILLIAM R.  
REGISTRATION NUMBER: 41,173

REFERENCE/DOCKET NUMBER: BB-1098-A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 302-992-4926  
TELEFAX: 302-773-0164  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 327 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-154-874-8

Query Match 1.4%; Score 7; DB 3; Length 327;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 343 GAGILK 349  
DB 107 GAGILK 113

RESULT 23  
US-08-931-668-8  
Sequence 8, Application US/08931668  
Patent No. 6521433  
GENERAL INFORMATION:  
APPLICANT: FADER, GARY M.  
TITLE OF INVENTION: CDNA SEQUENCES FROM SOYBEAN THAT ENCODE  
TITLE OF INVENTION: ACTIVITIES ASSOCIATED WITH ISOPOLYMER  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY  
STREET: 1007 MARKET STREET  
CITY: WILMINGTON  
STATE: DELAWARE  
COUNTRY: UNITED STATES OF AMERICA  
ZIP: 19898  
COMPUTER READABLE FORM:  
MEDIUM TYPE: DISKETTE, 3.50 INCH  
COMPUTER: IBM PC COMPATIBLE  
OPERATING SYSTEM: MICROSOFT WORD FOR WINDOWS 95  
SOFTWARE: MICROSOFT WORD VERSION 7.0A  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/931,668  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: MAJARIAN, WILLIAM R.  
REGISTRATION NUMBER: 41,173  
REFERENCE/DOCKET NUMBER: BB-1098  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 302-992-4926  
TELEFAX: 302-773-0164  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 327 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-931-668-8

Query Match 1.4%; Score 7; DB 4; Length 327;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 343 GAGILK 349  
DB 107 GAGILK 113

RESULT 24  
US-09-252-991A-21843

; Sequence 21843, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 21843  
; LENGTH: 376  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-21843

Query Match 1.4%; Score 7; DB 4; Length 376;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 264 KVGTVR 270  
Db 322 KVGTVR 328

RESULT 25  
US-08-886-886-2  
; Sequence 2, Application US/0886886  
; Patent No. 6107068  
; GENERAL INFORMATION:  
; APPLICANT: Katz, Leonard  
; APPLICANT: Delcardayre, Stephen B.  
; APPLICANT: Davies, Julian E.  
; TITLE OF INVENTION: COENZYME A DISULFIDE REDUCTASE,  
; TITLE OF INVENTION: AND INHIBITORS THEREOF USEFUL AS ANTIMICROBIAL  
; TITLE OF INVENTION: AGENTS  
; NUMBER OF SEQUENCES: 40  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Abbott Laboratories  
; STREET: 100 Abbott Park Road  
; CITY: Abbott Park  
; STATE: IL  
; COUNTRY: USA  
; ZIP: 60064-3500  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: PASTESEQ for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/886,886  
; FILING DATE: 02-JUL-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Casuto, Dianne  
; REGISTRATION NUMBER: 40,943  
; REFERENCE/DOCKET NUMBER: 6016.US.P1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 847-938-3137  
; TELEFAX: 847-938-2623  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 438 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single

; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
US-08-886-886-2

Query Match 1.4%; Score 7; DB 3; Length 438;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 436 RAASIVA 442  
Db 300 RAASIVA 306

RESULT 26  
US-09-044-718-2  
; Sequence 2, Application US/09044718  
; Patent No. 6391605  
; GENERAL INFORMATION:  
; APPLICANT: KOSTREMA, Dirk  
; APPLICANT: PASAMONTES, Luis  
; APPLICANT: TOMSCHY, Andrea  
; APPLICANT: van LOON, Adolphus  
; APPLICANT: VOGEL, Kurt  
; APPLICANT: WYSS, Markus  
; TITLE OF INVENTION: MODIFIED PHYTASES  
; FILE REFERENCE: Modified Phytases  
; CURRENT APPLICATION NUMBER: US/09/044,718  
; CURRENT FILING DATE: 1998-03-19  
; PRIOR APPLICATION NUMBER: EP 97810175.6  
; PRIOR FILING DATE: 1997-03-25  
; NUMBER OF SEQ ID NOS: 82  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 438  
; TYPE: PRT  
; ORGANISM: Aspergillus terreus  
US-09-044-718-2

Query Match 1.4%; Score 7; DB 4; Length 438;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 487 QLSADDV 493  
Db 223 QLSADDV 229

RESULT 27  
US-09-296-284-26  
; Sequence 26, Application US/09296284A  
; Patent No. 6204040  
; GENERAL INFORMATION:  
; APPLICANT: Choi, Eun-Sung  
; APPLICANT: Rhee, Sang-Ki  
; APPLICANT: Lee, Eun-Hae  
; TITLE OF INVENTION: Glucanobacter Suboxydans Sorbitol Dehydrogenase, Genes  
; TITLE OF INVENTION: and Methods of Use Thereof  
; FILE REFERENCE: 1533.0870000  
; CURRENT APPLICATION NUMBER: US/09/296,284A  
; CURRENT FILING DATE: 1999-04-22  
; NUMBER OF SEQ ID NOS: 87  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 26  
; LENGTH: 442  
; TYPE: PRT  
; ORGANISM: Glucanobacter suboxydans  
US-09-296-284-26

Query Match 1.4%; Score 7; DB 3; Length 442;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 465 LSDQOIA 471

Db 365 LSDQIA 371

RESULT 28  
US-09-044-718-15  
Sequence 15, Application US/09044718  
Patent No. 6391605  
GENERAL INFORMATION:  
APPLICANT: KOSTREMA, Dirk  
APPLICANT: PASAMONTES, Luis  
APPLICANT: TOMSCHY, Andrea  
APPLICANT: van LOON, Adolphus  
APPLICANT: VOGEL, Kurt  
APPLICANT: WYSS, Markus  
TITLE OF INVENTION: MODIFIED PHYTASES  
FILE REFERENCE: Modified Phytases  
CURRENT APPLICATION NUMBER: US/09/044,718  
CURRENT FILING DATE: 1998-03-19  
PRIOR APPLICATION NUMBER: EP 97810175.6  
PRIOR FILING DATE: 1997-03-25  
NUMBER OF SEQ ID NOS: 82  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 15  
LENGTH: 450  
TYPE: PRT  
ORGANISM: Aspergillus terreus  
US-09-044-718-15

Query Match 1.4%; Score 7; DB 4; Length 450;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 487 QLSADV 493  
Db 235 QLSADV 241

RESULT 29  
US-08-868-435-35  
Sequence 35, Application US/08868435  
Patent No. 6291221  
GENERAL INFORMATION:  
APPLICANT: Van Loon, Adolphus  
APPLICANT: Mitchell, David  
TITLE OF INVENTION: POLYPEPTIDES WITH PHYTASE ACTIVITY  
NUMBER OF SEQUENCES: 35  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hoffmann-La Roche Inc.  
STREET: 340 Kingsland Street  
City: Nutley  
STATE: New Jersey  
COUNTRY: United States of America  
ZIP: 07110  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/868,435  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/744,231  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Kaas, Alan P  
REGISTRATION NUMBER: 32142  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (201) 235-4205  
TELEFAX: (201) 235-2363

INFORMATION FOR SEQ ID NO: 35:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 466 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 27  
OTHER INFORMATION: /note="potential N-glycosylation site"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 120  
OTHER INFORMATION: /note="potential N-glycosylation site"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 207  
OTHER INFORMATION: /note="potential N-glycosylation site"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 230  
OTHER INFORMATION: /note="potential N-glycosylation site"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 352  
OTHER INFORMATION: /note="potential N-glycosylation site"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 376  
OTHER INFORMATION: /note="potential N-glycosylation site"  
US-08-868-435-35

Query Match 1.4%; Score 7; DB 3; Length 466;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 487 QLSADV 493  
Db 251 QLSADV 257

RESULT 30  
US-08-744-231-35  
Sequence 35, Application US/08744231  
Patent No. 6358722  
GENERAL INFORMATION:  
APPLICANT: Van Loon, Adolphus  
APPLICANT: Mitchell, David  
TITLE OF INVENTION: POLYPEPTIDES WITH PHYTASE ACTIVITY  
NUMBER OF SEQUENCES: 35  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hoffmann-La Roche Inc.  
STREET: 340 Kingsland Street  
City: Nutley  
STATE: New Jersey  
COUNTRY: United States of America  
ZIP: 07110  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/744,231  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/424,757  
FILING DATE: 18-Apr-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Kaas, Alan P  
REGISTRATION NUMBER: 32142  
REFERENCE/DOCKET NUMBER: Case Docket 9339

```
TELECOMMUNICATION INFORMATION:
TELEPHONE: (201) 235-4205
TELEFAX: (201) 235-2363
INFORMATION FOR SEQ ID NO: 35:
SEQUENCE CHARACTERISTICS:
LENGTH: 466 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
FEATURE:
NAME/KEY: misc_feature
LOCATION: 27
OTHER INFORMATION: /note="potential N-glycosylation site"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 120
OTHER INFORMATION: /note="potential N-glycosylation site"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 207
OTHER INFORMATION: /note="potential N-glycosylation site"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 230
OTHER INFORMATION: /note="potential N-glycosylation site"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 352
OTHER INFORMATION: /note="potential N-glycosylation site"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 376
OTHER INFORMATION: /note="potential N-glycosylation site"
OTHER INFORMATION: /note="potential N-glycosylation site"
US-08-744-231-35

Query Match
Best Local Similarity 1.4%; Score 7; DB 4; Length 466;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 487 QLSADDV 493
Db 251 QLSADDV 257

RESULT 31
US-09-296-284-5
Sequence 5, Application US/09296284A
Patent No. 6204040
GENERAL INFORMATION:
APPLICANT: Choi, Eun-Sung
APPLICANT: Rhee, Sang-Ki
APPLICANT: Lee, Eun-Hae
TITLE OF INVENTION: Glucanobacter Suboxydans Sorbitol Dehydrogenase, Genes
FILE REFERENCE: 1533.087000
CURRENT APPLICATION NUMBER: US/09/296,284A
CURRENT FILING DATE: 1999-04-22
NUMBER OF SEQ ID NOS: 87
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 5
LENGTH: 478
TYPE: PRT
ORGANISM: Glucanobacter suboxydans
US-09-296-284-5

Query Match
Best Local Similarity 1.4%; Score 7; DB 3; Length 478;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 465 LSDOOIA 471
Db 401 LSDOOIA 407
```

```
RESULT 32
US-08-836-620A-18
Sequence 18, Application US/08836620A
Patent No. 5958710
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: Orphan receptor
NUMBER OF SEQUENCES: 19
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/836,620A
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/EP96/03933
FILING DATE:
APPLICATION NUMBER: GB 9518272.1
FILING DATE: 08-SEP-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9605550.4
FILING DATE: 15-MAR-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9607532.0
FILING DATE: 11-APR-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9609576.5
FILING DATE: 08-MAY-1996
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 518 amino acids
TYPE: amino acid
TOPOLOGY: linear
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
US-08-836-620A-18

Query Match
Best Local Similarity 1.4%; Score 7; DB 2; Length 518;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 148 GPGGAE 154
Db 466 GPGGAE 472

RESULT 33
US-08-878-563A-3
Sequence 3, Application US/08878563A
Patent No. 5851674
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
APPLICANT: Lal, Preeti
APPLICANT: Shah, Purvi
TITLE OF INVENTION: INSULIN RECEPTOR TYROSINE KINASE SUBSTRATE
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSER: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/878,563A
```

FILING DATE: Filed Herewith  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0323 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 521 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: 1203820  
US-08-878-563A-3

Query\*Match 1.4%; Score 7; DB 2; Length 521;  
Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 29 KAAQPKS 35  
|||||  
Db 318 KAAQPKS 324

RESULT 34  
US-09-270-117-3  
Sequence 3, Application US/09270117  
Patent No. 6265550  
GENERAL INFORMATION:  
APPLICANT: Hillman, Jennifer L.  
APPLICANT: Lal, Preeti  
APPLICANT: Shah, Purni  
TITLE OF INVENTION: INSULIN RECEPTOR TYROSINE KINASE SUBSTRATE  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/270.117  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/878,563  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0323 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 521 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:

LIBRARY: GenBank  
CLONE: 1203820  
US-09-270-117-3

Query\*Match 1.4%; Score 7; DB 3; Length 521;  
Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 29 KAAQPKS 35  
|||||  
Db 318 KAAQPKS 324

RESULT 35  
US-08-878-563A-1  
Sequence 1, Application US/08878563A  
Patent No. 5891674  
GENERAL INFORMATION:  
APPLICANT: Hillman, Jennifer L.  
APPLICANT: Lal, Preeti  
APPLICANT: Shah, Purni  
TITLE OF INVENTION: INSULIN RECEPTOR TYROSINE KINASE SUBSTRATE  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/878,563A  
FILING DATE: Filed Herewith  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0323 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 534 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: BRSTNOT04  
CLONE: 918158  
US-08-878-563A-1

Query\*Match 1.4%; Score 7; DB 2; Length 534;  
Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 29 KAAQPKS 35  
|||||  
Db 317 KAAQPKS 323

RESULT 36  
US-09-270-117-1  
Sequence 1, Application US/09270117  
Patent No. 6265550  
GENERAL INFORMATION:



APPLICANT: Hillman, Jennifer L.  
APPLICANT: Lal, Preeti  
APPLICANT: Shah, Purni  
TITLE OF INVENTION: INSULIN RECEPTOR TYROSINE KINASE SUBSTRATE  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/270,117  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/878,563  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0323 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 534 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: BRSTNOT04  
CLONE: 918158  
US-09-270-117-1

Query Match 1.4%; Score 7; DB 3; Length 534;  
Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 29 KAAPKS 35  
Db 317 KAAPKS 323  
RESULT 37  
US-09-328-352-5771  
Sequence 5771, Application US/09328352  
Patent No. 6562958  
GENERAL INFORMATION:  
APPLICANT: Gary L. Breton et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
FILE REFERENCE: GTC99-03PA  
CURRENT APPLICATION NUMBER: US/09/328,352  
CURRENT FILING DATE: 1999-06-04  
NUMBER OF SEQ ID NOS: 8252  
SEQ ID NO 5771  
LENGTH: 539  
TYPE: PRT  
ORGANISM: Acinetobacter baumannii  
US-09-328-352-5771

Query Match 1.4%; Score 7; DB 4; Length 539;  
Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 257 GENALKA 263

Db 519 GENALKA 525  
RESULT 38  
US-09-252-991A-20407  
Sequence 20407, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 20407  
LENGTH: 580  
TYPE: PRT  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-20407

Query Match 1.4%; Score 7; DB 4; Length 580;  
Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 370 PEGAPQA 376  
Db 466 PEGAPQA 472

RESULT 39  
US-09-252-991A-19164  
Sequence 19164, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 19164  
LENGTH: 616  
TYPE: PRT  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-19164

Query Match 1.4%; Score 7; DB 4; Length 616;  
Best Local Similarity 100.0%; Pred. No. 2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 96 RLADGVE 102  
Db 432 RLADGVE 438

RESULT 40  
US-09-252-991A-26100  
Sequence 26100, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252.991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ. ID NOS: 33142  
; SEQ. ID NO 26100  
; LENGTH: 727  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-26100

Query Match 1.4%; Score 7; DB 4; Length 727;  
Best Local Similarity 100.0%; Pred. No. 2.3e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 96 RLADGVE 102  
|||  
|||  
Db 637 RLADGVE 643

Search completed: August 27, 2003, 18:46:11  
Job time : 41 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: August 27, 2003, 18:45:18 ; Search time 60 seconds

(without alignments)  
1144.408 Million cell updates/sec

Title: US-10-088-045-2

Perfect score: 502  
Sequence: 1 MSKPTLKITTLICALSALML.....NKGQGLSADVAKKTKPEN 502Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 510680 seqs, 136781880 residues

Word size : 0

Total number of hits satisfying chosen parameters: 510680

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database : Published Applications AA:\*

1: /cgn2\_6/ptodata/2/pubppaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/2/pubppaa/PCT\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/2/pubppaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/2/pubppaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/2/pubppaa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/2/pubppaa/PCTUS\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/2/pubppaa/US08\_NEW\_PUB.pep:\*  
8: /cgn2\_6/ptodata/2/pubppaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/2/pubppaa/US09A\_PUBCOMB.pep:\*  
10: /cgn2\_6/ptodata/2/pubppaa/US09B\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/2/pubppaa/US09C\_PUBCOMB.pep:\*  
12: /cgn2\_6/ptodata/2/pubppaa/US09\_NEW\_PUB.pep:\*  
13: /cgn2\_6/ptodata/2/pubppaa/US10A\_PUBCOMB.pep:\*  
14: /cgn2\_6/ptodata/2/pubppaa/US10B\_PUBCOMB.pep:\*  
15: /cgn2\_6/ptodata/2/pubppaa/US10C\_PUBCOMB.pep:\*  
16: /cgn2\_6/ptodata/2/pubppaa/US10\_NEW\_PUB.pep:\*  
17: /cgn2\_6/ptodata/2/pubppaa/US60\_NEW\_PUB.pep:\*  
18: /cgn2\_6/ptodata/2/pubppaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7	1.4	39	12	US-09-890-688-140
2	7	1.4	71	9	US-09-864-761-44403
3	7	1.4	104	15	US-10-026-741-100
4	7	1.4	156	15	US-10-101-464A-509
5	7	1.4	157	15	US-10-156-761-13774
6	7	1.4	180	14	US-10-027-450-28
7	7	1.4	233	14	US-10-138-787-7
8	7	1.4	238	9	US-09-904-954-2
9	7	1.4	238	12	US-10-241-220-72
10	7	1.4	278	10	US-09-738-626-5155
11	7	1.4	300	15	US-10-156-761-14202
12	7	1.4	312	10	US-09-967-347-11
13	7	1.4	411	12	US-10-092-947A-37
14	7	1.4	436	9	US-09-730-772-14
15	7	1.4	436	9	US-09-735-849-14

16	7	1.4	438	15	US-10-062-848-2	Sequence 2, Appli
17	7	1.4	439	10	US-09-925-637-20	Sequence 20, Appli
18	7	1.4	439	15	US-10-084-205-20	Sequence 10, Appli
19	7	1.4	450	15	US-10-062-848-15	Sequence 15, Appli
20	7	1.4	451	11	US-09-843-250-24	Sequence 24, Appli
21	7	1.4	518	15	US-10-278-481-18	Sequence 18, Appli
22	7	1.4	519	15	US-10-211-239-9	Sequence 1, Appli
23	7	1.4	549	15	US-10-156-761-14029	Sequence 14029, A
24	7	1.4	611	9	US-09-815-242-12111	Sequence 12111, A
25	7	1.4	724	15	US-10-156-761-9578	Sequence 9578, Ap
26	7	1.4	792	10	US-09-995-587A-11	Sequence 11, Appli
27	7	1.4	844	12	US-09-882-227-308	Sequence 308, App
28	7	1.4	859	8	US-08-945-749-1	Sequence 1, Appli
29	7	1.4	886	8	US-08-781-986A-5235	Sequence 5235, Ap
30	7	1.4	960	15	US-10-188-308-21	Sequence 21, Appli
31	7	1.4	960	15	US-10-188-296-21	Sequence 21, Appli
32	7	1.4	960	15	US-10-188-341-21	Sequence 21, Appli
33	7	1.4	960	15	US-10-188-297-21	Sequence 21, Appli
34	7	1.4	962	15	US-10-188-308-3	Sequence 3, Appli
35	7	1.4	962	15	US-10-188-308-24	Sequence 24, Appli
36	7	1.4	962	15	US-10-188-296-3	Sequence 3, Appli
37	7	1.4	962	15	US-10-188-296-24	Sequence 24, Appli
38	7	1.4	962	15	US-10-188-341-3	Sequence 3, Appli
39	7	1.4	962	15	US-10-188-341-24	Sequence 24, Appli
40	7	1.4	962	15	US-10-188-297-3	Sequence 3, Appli
41	7	1.4	962	15	US-10-188-297-24	Sequence 24, Appli
42	7	1.4	962	15	US-10-174-613-4	Sequence 4, Appli
43	7	1.4	962	15	US-10-174-613-5	Sequence 5, Appli
44	7	1.4	987	15	US-10-188-308-22	Sequence 22, Appli
45	7	1.4	987	15	US-10-188-296-22	Sequence 22, Appli
46	7	1.4	987	15	US-10-188-341-22	Sequence 22, Appli
47	7	1.4	987	15	US-10-188-297-22	Sequence 22, Appli
48	7	1.4	988	11	US-09-875-321-5	Sequence 5, Appli
49	7	1.4	988	11	US-09-875-321-12	Sequence 12, Appli
50	7	1.4	988	15	US-10-128-323-2	Sequence 2, Appli
51	7	1.4	988	15	US-10-162-012-5	Sequence 5, Appli
52	7	1.4	988	15	US-10-162-012-12	Sequence 12, Appli
53	7	1.4	988	15	US-10-174-613-2	Sequence 2, Appli
54	7	1.4	988	15	US-10-174-613-3	Sequence 3, Appli
55	7	1.4	989	15	US-10-188-308-4	Sequence 4, Appli
56	7	1.4	989	15	US-10-188-308-23	Sequence 23, Appli
57	7	1.4	989	15	US-10-188-296-4	Sequence 4, Appli
58	7	1.4	989	15	US-10-188-296-23	Sequence 23, Appli
59	7	1.4	989	15	US-10-188-341-4	Sequence 4, Appli
60	7	1.4	989	15	US-10-188-341-23	Sequence 23, Appli
61	7	1.4	989	15	US-10-188-297-4	Sequence 23, Appli
62	7	1.4	989	15	US-10-188-297-23	Sequence 23, Appli
63	7	1.4	989	15	US-10-174-613-6	Sequence 6, Appli
64	7	1.4	1095	15	US-10-128-714-3039	Sequence 3039, Ap
65	7	1.4	1174	15	US-10-174-613-7	Sequence 7, Appli
66	7	1.4	1277	15	US-10-128-714-8039	Sequence 8039, Ap
67	7	1.4	1676	15	US-10-128-714-8246	Sequence 8246, Ap
68	7	1.2	11	15	US-10-108-795-28	Sequence 28, Appli
69	6	1.2	11	15	US-10-108-795-29	Sequence 29, Appli
70	6	1.2	11	15	US-10-115-365-28	Sequence 28, Appli
71	6	1.2	11	15	US-10-115-365-29	Sequence 29, Appli
72	6	1.2	23	11	US-09-774-639-331	Sequence 331, App
73	6	1.2	23	11	US-09-969-720-263	Sequence 263, App
74	6	1.2	25	9	US-09-864-761-39462	Sequence 39462, A
75	6	1.2	34	9	US-09-880-578-36	Sequence 36, Appli
76	6	1.2	34	9	US-09-864-761-39670	Sequence 39670, A
77	6	1.2	34	9	US-09-864-761-42206	Sequence 42206, A
78	6	1.2	38	10	US-09-919-473-12	Sequence 12, Appli
79	6	1.2	38	15	US-10-099-766-6	Sequence 6, Appli
80	6	1.2	40	9	US-09-864-761-43181	Sequence 43181, A
81	6	1.2	40	9	US-09-682-706-15	Sequence 15, Appli
82	6	1.2	47	8	US-08-424-550B-556	Sequence 556, App
83	6	1.2	49	9	US-09-864-761-36011	Sequence 36011, A
84	6	1.2	57	10	US-09-738-626-5498	Sequence 5498, Ap
85	6	1.2	58	11	US-09-764-891-4535	Sequence 4535, Ap
86	6	1.2	59	9	US-09-864-761-41409	Sequence 41409, A
87	6	1.2	59	9	US-09-864-761-46850	Sequence 46850, A
88	6	1.2	60	15	US-10-179-784-8	Sequence 8, Appli













819	6	1.2	373	15	US-10-232-222-130	Sequence 130, App	892	6	1.2	385	15	US-10-245-165-84	Sequence 84, Appl
820	6	1.2	373	15	US-10-166-709A-59	Sequence 59, Appl	893	6	1.2	385	15	US-10-245-427-84	Sequence 84, Appl
821	6	1.2	373	15	US-10-192-007-388	Sequence 388, App	894	6	1.2	385	15	US-10-245-473-84	Sequence 84, Appl
822	6	1.2	373	15	US-10-194-359-388	Sequence 388, App	895	6	1.2	385	15	US-10-245-770-84	Sequence 84, Appl
823	6	1.2	373	15	US-10-223-084-64	Sequence 64, Appl	896	6	1.2	385	15	US-10-245-877-84	Sequence 84, Appl
824	6	1.2	373	15	US-10-223-088-64	Sequence 64, Appl	897	6	1.2	385	15	US-10-245-976-84	Sequence 84, Appl
825	6	1.2	373	15	US-10-223-090-64	Sequence 64, Appl	898	6	1.2	385	15	US-10-245-930-84	Sequence 84, Appl
826	6	1.2	373	15	US-10-219-070-110	Sequence 130, App	899	6	1.2	385	15	US-10-245-943-16	Sequence 16, Appl
827	6	1.2	373	15	US-10-219-472-130	Sequence 130, App	900	6	1.2	385	15	US-10-245-885-84	Sequence 84, Appl
828	6	1.2	373	15	US-10-219-527-130	Sequence 130, App	901	6	1.2	385	15	US-10-245-843-84	Sequence 84, Appl
829	6	1.2	373	15	US-10-227-877-130	Sequence 130, App	902	6	1.2	385	15	US-10-237-636-84	Sequence 84, Appl
830	6	1.2	373	15	US-10-223-087-64	Sequence 64, Appl	903	6	1.2	385	15	US-10-238-335-84	Sequence 84, Appl
831	6	1.2	373	15	US-10-127-847A-388	Sequence 388, App	904	6	1.2	385	15	US-10-238-336-84	Sequence 84, Appl
832	6	1.2	373	15	US-10-223-083-64	Sequence 64, Appl	905	6	1.2	385	15	US-10-238-411-84	Sequence 84, Appl
833	6	1.2	373	15	US-10-216-166-130	Sequence 130, App	906	6	1.2	385	15	US-10-243-124-84	Sequence 84, Appl
834	6	1.2	373	15	US-10-218-612-130	Sequence 130, App	907	6	1.2	385	15	US-10-243-425-84	Sequence 84, Appl
835	6	1.2	373	15	US-10-175-590-388	Sequence 388, App	908	6	1.2	385	15	US-10-243-446-84	Sequence 84, Appl
836	6	1.2	373	15	US-10-223-089-64	Sequence 64, Appl	909	6	1.2	385	15	US-10-245-874-84	Sequence 84, Appl
837	6	1.2	373	16	US-10-137-866-388	Sequence 388, App	910	6	1.2	385	15	US-10-245-653-84	Sequence 84, Appl
838	6	1.2	373	16	US-10-146-726-388	Sequence 388, App	911	6	1.2	385	15	US-10-245-178-84	Sequence 84, Appl
839	6	1.2	373	16	US-10-146-727-388	Sequence 388, App	912	6	1.2	385	15	US-10-243-388-84	Sequence 84, Appl
840	6	1.2	373	16	US-10-146-788-388	Sequence 388, App	913	6	1.2	385	15	US-10-244-947-84	Sequence 84, Appl
841	6	1.2	373	16	US-10-152-380-388	Sequence 388, App	914	6	1.2	385	15	US-10-244-968-84	Sequence 84, Appl
842	6	1.2	373	16	US-10-153-934-388	Sequence 388, App	915	6	1.2	385	15	US-10-244-990-84	Sequence 84, Appl
843	6	1.2	374	9	US-09-688-615-61	Sequence 61, Appl	916	6	1.2	385	15	US-10-245-079-84	Sequence 84, Appl
844	6	1.2	374	15	US-10-127-032-116	Sequence 116, App	917	6	1.2	385	15	US-10-245-127-84	Sequence 84, Appl
845	6	1.2	375	10	US-09-978-249-10	Sequence 7, Appl	918	6	1.2	385	15	US-10-245-207-84	Sequence 84, Appl
846	6	1.2	377	15	US-10-149-819-7	Sequence 22, Appl	919	6	1.2	385	15	US-10-245-646-84	Sequence 84, Appl
847	6	1.2	377	15	US-10-318-142-22	Sequence 13402, A	920	6	1.2	385	15	US-10-245-695-84	Sequence 84, Appl
848	6	1.2	378	9	US-09-615-242-13402	Sequence 9745, Ap	921	6	1.2	385	15	US-10-245-659-84	Sequence 84, Appl
849	6	1.2	378	15	US-10-156-761-9745	Sequence 6, Appl	922	6	1.2	385	15	US-10-245-777-84	Sequence 84, Appl
850	6	1.2	379	11	US-09-813-408-6	Sequence 8, Appl	923	6	1.2	385	15	US-10-245-878-84	Sequence 84, Appl
851	6	1.2	380	10	US-09-755-665-8	Sequence 82, Appl	924	6	1.2	385	15	US-10-245-899-84	Sequence 84, Appl
852	6	1.2	380	11	US-09-955-999-82	Sequence 13, Appl	925	6	1.2	385	15	US-10-245-900-84	Sequence 84, Appl
853	6	1.2	381	10	US-09-964-899-13	Sequence 46, Appl	926	6	1.2	385	15	US-10-245-908-84	Sequence 84, Appl
854	6	1.2	381	15	US-10-161-572-46	Sequence 7, Appl	927	6	1.2	385	15	US-10-245-927-84	Sequence 84, Appl
855	6	1.2	383	9	US-09-939-484-7	Sequence 7, Appl	928	6	1.2	385	15	US-10-245-454-84	Sequence 84, Appl
856	6	1.2	384	9	US-09-939-483-7	Sequence 6407, Ap	929	6	1.2	385	15	US-10-238-261-84	Sequence 84, Appl
857	6	1.2	384	10	US-09-738-626-6407	Sequence 13, Appl	930	6	1.2	385	15	US-10-238-324-84	Sequence 84, Appl
858	6	1.2	384	14	US-10-077-111-13	Sequence 641, App	931	6	1.2	385	15	US-10-238-324-84	Sequence 84, Appl
859	6	1.2	384	15	US-10-225-567A-641	Sequence 20, Appl	932	6	1.2	385	15	US-10-241-860-84	Sequence 84, Appl
860	6	1.2	385	9	US-09-680-578-20	Sequence 84, Appl	933	6	1.2	385	15	US-10-242-172-84	Sequence 84, Appl
861	6	1.2	385	12	US-10-237-496-84	Sequence 84, Appl	934	6	1.2	385	15	US-10-242-652-84	Sequence 84, Appl
862	6	1.2	385	12	US-10-242-074-84	Sequence 84, Appl	935	6	1.2	385	15	US-10-242-990-84	Sequence 84, Appl
863	6	1.2	385	12	US-10-242-505-84	Sequence 84, Appl	936	6	1.2	385	15	US-10-243-023-84	Sequence 84, Appl
864	6	1.2	385	12	US-10-242-574-84	Sequence 84, Appl	937	6	1.2	385	15	US-10-243-103-84	Sequence 84, Appl
865	6	1.2	385	12	US-10-243-264-84	Sequence 84, Appl	938	6	1.2	385	15	US-10-243-276-84	Sequence 84, Appl
866	6	1.2	385	12	US-10-243-282-84	Sequence 84, Appl	939	6	1.2	385	15	US-10-243-326-84	Sequence 84, Appl
867	6	1.2	385	12	US-10-243-402-84	Sequence 84, Appl	940	6	1.2	385	15	US-10-243-364-84	Sequence 84, Appl
868	6	1.2	385	12	US-10-243-431-84	Sequence 84, Appl	941	6	1.2	385	15	US-10-243-494-84	Sequence 84, Appl
869	6	1.2	385	12	US-10-245-164-84	Sequence 84, Appl	942	6	1.2	385	15	US-10-244-995-84	Sequence 84, Appl
870	6	1.2	385	15	US-10-245-103-84	Sequence 84, Appl	943	6	1.2	385	15	US-10-245-220-84	Sequence 84, Appl
871	6	1.2	385	15	US-10-245-107-84	Sequence 84, Appl	944	6	1.2	385	15	US-10-245-253-84	Sequence 84, Appl
872	6	1.2	385	15	US-10-245-143-84	Sequence 84, Appl	945	6	1.2	385	15	US-10-245-479-84	Sequence 84, Appl
873	6	1.2	385	15	US-10-245-771-84	Sequence 84, Appl	946	6	1.2	385	15	US-10-245-499-84	Sequence 84, Appl
874	6	1.2	385	15	US-10-245-851-84	Sequence 84, Appl	947	6	1.2	385	15	US-10-245-772-84	Sequence 84, Appl
875	6	1.2	385	15	US-10-245-883-84	Sequence 84, Appl	948	6	1.2	385	15	US-10-245-811-84	Sequence 84, Appl
876	6	1.2	385	15	US-10-237-533-84	Sequence 84, Appl	949	6	1.2	385	15	US-10-245-812-84	Sequence 84, Appl
877	6	1.2	385	15	US-10-238-183-84	Sequence 84, Appl	950	6	1.2	385	15	US-10-245-852-84	Sequence 84, Appl
878	6	1.2	385	15	US-10-238-283-84	Sequence 84, Appl	951	6	1.2	385	15	US-10-245-875-84	Sequence 84, Appl
879	6	1.2	385	15	US-10-238-370-84	Sequence 84, Appl	952	6	1.2	385	15	US-10-245-881-84	Sequence 84, Appl
880	6	1.2	385	15	US-10-245-055-84	Sequence 84, Appl	953	6	1.2	385	15	US-10-245-911-84	Sequence 84, Appl
881	6	1.2	385	15	US-10-245-147-84	Sequence 84, Appl	954	6	1.2	385	15	US-10-245-913-84	Sequence 84, Appl
882	6	1.2	385	15	US-10-245-730-84	Sequence 84, Appl	955	6	1.2	385	15	US-10-246-080-84	Sequence 84, Appl
883	6	1.2	385	15	US-10-245-733-84	Sequence 84, Appl	956	6	1.2	385	15	US-10-246-121-84	Sequence 84, Appl
884	6	1.2	385	15	US-10-246-210-84	Sequence 84, Appl	957	6	1.2	385	15	US-10-246-305-84	Sequence 84, Appl
885	6	1.2	385	15	US-10-239-196-84	Sequence 84, Appl	958	6	1.2	385	15	US-10-246-329-84	Sequence 84, Appl
886	6	1.2	385	15	US-10-243-024-84	Sequence 84, Appl	959	6	1.2	385	15	US-10-247-036-84	Sequence 84, Appl
887	6	1.2	385	15	US-10-243-409-84	Sequence 84, Appl	960	6	1.2	385	15	US-10-243-255-84	Sequence 84, Appl
888	6	1.2	385	15	US-10-245-621-84	Sequence 84, Appl	961	6	1.2	385	15	US-10-243-910-84	Sequence 84, Appl
889	6	1.2	385	15	US-10-245-880-84	Sequence 84, Appl	962	6	1.2	385	15	US-10-245-910-84	Sequence 84, Appl
890	6	1.2	385	15	US-10-245-033-84	Sequence 84, Appl	963	6	1.2	385	15	US-10-246-098-84	Sequence 84, Appl
891	6	1.2	385	15	US-10-243-095-84	Sequence 84, Appl	964	6	1.2	386	10	US-09-753-143-75	Sequence 75, Appl

965 6 1.2 386 15 US-10-034-937-2 Sequence 2, Appl  
966 6 1.2 386 15 US-10-122-706-2 Sequence 2, Appl  
967 6 1.2 387 9 US-09-866-987-9 Sequence 9, Appl  
968 6 1.2 387 15 US-10-034-937-42 Sequence 42, Appl  
969 6 1.2 387 15 US-10-156-761-14269 Sequence 14269, A  
970 6 1.2 387 15 US-10-156-761-14583 Sequence 14583, A  
971 6 1.2 388 9 US-09-880-578-17 Sequence 17, Appl  
972 6 1.2 388 15 US-10-161-572-47 Sequence 47, Appl  
973 6 1.2 388 15 US-10-230-318-1 Sequence 1, Appl  
974 6 1.2 389 9 US-09-880-578-22 Sequence 22, Appl  
975 6 1.2 389 9 US-09-880-578-24 Sequence 24, Appl  
976 6 1.2 389 9 US-09-880-578-25 Sequence 25, Appl  
977 6 1.2 389 9 US-09-880-578-26 Sequence 26, Appl  
978 6 1.2 389 9 US-09-880-578-27 Sequence 27, Appl  
979 6 1.2 389 9 US-09-880-578-28 Sequence 28, Appl  
980 6 1.2 389 9 US-09-880-578-29 Sequence 29, Appl  
981 6 1.2 389 9 US-09-880-578-30 Sequence 30, Appl  
982 6 1.2 389 9 US-09-880-578-31 Sequence 31, Appl  
983 6 1.2 390 10 US-09-939-521-10 Sequence 10, Appl  
984 6 1.2 390 15 US-10-034-937-4 Sequence 4, Appl  
985 6 1.2 390 15 US-10-034-937-6 Sequence 6, Appl  
986 6 1.2 390 15 US-10-034-937-8 Sequence 8, Appl  
987 6 1.2 390 15 US-10-034-937-10 Sequence 10, Appl  
988 6 1.2 390 15 US-10-034-937-12 Sequence 12, Appl  
989 6 1.2 390 15 US-10-034-937-14 Sequence 14, Appl  
990 6 1.2 390 15 US-10-034-937-16 Sequence 16, Appl  
991 6 1.2 390 15 US-10-034-937-18 Sequence 18, Appl  
992 6 1.2 390 15 US-10-034-937-20 Sequence 20, Appl  
993 6 1.2 390 15 US-10-034-937-22 Sequence 22, Appl  
994 6 1.2 390 15 US-10-034-937-24 Sequence 24, Appl  
995 6 1.2 390 15 US-10-034-937-26 Sequence 26, Appl  
996 6 1.2 390 15 US-10-034-937-28 Sequence 28, Appl  
997 6 1.2 390 15 US-10-034-937-30 Sequence 30, Appl  
998 6 1.2 390 15 US-10-034-937-32 Sequence 32, Appl  
999 6 1.2 390 15 US-10-034-937-34 Sequence 34, Appl  
1000 6 1.2 390 15 US-10-034-937-36 Sequence 36, Appl

## ALIGNMENTS

RESULT 1  
US-09-890-688-140  
; Sequence 140, Application US/09890688  
; Publication No. US2003014447SA1  
; GENERAL INFORMATION:  
; APPLICANT: Seishi KATO  
; APPLICANT: Chikashi EGUCHI  
; APPLICANT: Mihoro SAKI  
; TITLE OF INVENTION: Human Proteins and cDNAs thereof  
; FILE REFERENCE: 2001-1102A/WMC/00653  
; CURRENT APPLICATION NUMBER: US/09/890,688  
; PRIOR FILING DATE: 2001-08-06  
; PRIOR APPLICATION NUMBER: JP 11-346863  
; PRIOR FILING DATE: 1999-12-06  
; PRIOR APPLICATION NUMBER: JP 11-34684  
; PRIOR FILING DATE: 1999-12-06  
; PRIOR APPLICATION NUMBER: JP 2000-31062  
; PRIOR FILING DATE: 2000-02-08  
; PRIOR APPLICATION NUMBER: JP 2000-34091  
; PRIOR FILING DATE: 2000-02-10  
; PRIOR APPLICATION NUMBER: JP 2000-34090  
; PRIOR FILING DATE: 2000-02-10  
; PRIOR APPLICATION NUMBER: JP 2000-35829  
; PRIOR FILING DATE: 2000-02-14  
; PRIOR APPLICATION NUMBER: JP 2000-35839  
; PRIOR FILING DATE: 2000-02-14  
; PRIOR APPLICATION NUMBER: JP 2000-71161  
; PRIOR FILING DATE: 2000-03-14  
; PRIOR APPLICATION NUMBER: JP 2000-160851  
; PRIOR FILING DATE: 2000-05-10  
; NUMBER OF SEQ ID NOS: 160  
; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 140  
; LENGTH: 39  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-890-688-140  
Query Match 1.4%; Score 7; DB 12; Length 39;  
Best Local Similarity 100.0%; Pred. No. 33;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 150 GCGAENS 156  
Db 6 GCGAENS 12  
RESULT 2  
US-09-864-761-44403  
; Sequence 44403, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharon G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; FILE REFERENCE: Aecmics-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774,203  
; PRIOR FILING DATE: 2001-01-29  
; NUMBER OF SEQ ID NOS: 49117  
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 44403  
; LENGTH: 71  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AC005015.2

OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1  
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.87  
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.78  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.8  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.5  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.78  
OTHER INFORMATION: EST HUMAN HIT: AA077633.1, VALUE 2.00e-19  
OTHER INFORMATION: SWISSPROT HIT: O42184, VALUE 3.00e-11  
US-09-864-761-44403

Query Match  
Best Local Similarity 1.4%; Score 7; DB 9; Length 71;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 494 AKAKTK 500  
DB 28 AKAKTK 34

RESULT 3  
US-10-026-741-100  
Sequence 100, Application US/10026741  
Publication No. US20030049604A1  
GENERAL INFORMATION:  
APPLICANT: CHARNEAU, PIERRE  
CLAVEL, FRANCOISE  
BORMAN, ANDREW  
OUILLENT, CAROLINE  
GUETARD, DENISE  
MONTAGNIER, LUC  
DONJON DE SAINT-MARTIN, JACQUELINE  
COHEN, JACQUES

TITLE OF INVENTION: NUCLEOTIDE SEQUENCES OF HIV-1 TYPE (OR  
SUBTYPE) ANTIGENS

NUMBER OF SEQUENCES: 103

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &  
Dunnet, L.L.P.

STREET: 1300 I Street, N.W.

CITY: Washington

STATE: DC

COUNTRY: USA

ZIP: 20005-3315

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/026,741

FILING DATE: 27-Dec-2001

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/817,441

FILING DATE: 31-AUG-1998

APPLICATION NUMBER: PCT/FR 95/01391

FILING DATE: 20-OCT-1995

APPLICATION NUMBER: FR 9412554

FILING DATE: 20-OCT-1994

APPLICATION NUMBER: FR 9502526

FILING DATE: 03-MAR-1995

ATTORNEY/AGENT INFORMATION:

NAME: Meyers, Kenneth J.

REGISTRATION NUMBER: 25,146

REFERENCE/DOCKET NUMBER: 03260.6005-00000

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-408-4000

TELEFAX: 202-408-4400

INFORMATION FOR SEQ ID NO: 100:

SEQUENCE CHARACTERISTICS:

LENGTH: 104 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide  
SEQUENCE DESCRIPTION: SEQ ID NO: 100:  
US-10-026-741-100

Query Match  
Best Local Similarity 1.4%; Score 7; DB 15; Length 104;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 467 DQDIANY 473  
DB 83 DQDIANY 89

RESULT 4  
US-10-101-464A-509

Sequence 509, Application US/10101464A  
Publication No. US20030046728A1  
GENERAL INFORMATION:

APPLICANT: Strabala, Timothy  
APPLICANT: Nieuwenhuizen, Nicolaas

APPLICANT: Higgins, Colleen M.

TITLE OF INVENTION: Compositions Isolated from Plant Cells  
and Their Use in the Modification of Plant Cell Signaling

FILE REFERENCE: 11000.1020c2

CURRENT APPLICATION NUMBER: US/10/101,464A

CURRENT FILING DATE: 2002-03-18

PRIOR APPLICATION NUMBER: 09/704,302

PRIOR FILING DATE: 2000-11-01

PRIOR APPLICATION NUMBER: 09/228,986

PRIOR FILING DATE: 1999-01-12

PRIOR APPLICATION NUMBER: 60/162,866

PRIOR FILING DATE: 1999-11-01

PRIOR APPLICATION NUMBER: PCT/US00/00724

PRIOR FILING DATE: 2000-01-11

NUMBER OF SEQ ID NOS: 989

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 509

LENGTH: 156

TYPE: PRT

ORGANISM: Eucalyptus grandis

US-10-101-464A-509

Query Match  
Best Local Similarity 1.4%; Score 7; DB 15; Length 156;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 443 NGISGKI 449  
DB 115 NGISGKI 121

RESULT 5  
US-10-156-761-13774

Sequence 13774, Application US/10156761  
Publication No. US20030119018A1  
GENERAL INFORMATION:

APPLICANT: OKURA, SATOSHI

APPLICANT: IKEDA, HARUO

APPLICANT: ISHIKAWA, JUN

APPLICANT: HORIKAWA, HIROSHI

APPLICANT: SHIBA, TADAYOSHI

APPLICANT: SAKAKI, YOSHIYUKI

APPLICANT: HATTORI, MASAHIRA

TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES

FILE REFERENCE: 249-262

CURRENT APPLICATION NUMBER: US/10/156,761

CURRENT FILING DATE: 2002-05-29

PRIOR APPLICATION NUMBER: JP 2001-204089

PRIOR FILING DATE: 2001-05-30

PRIOR APPLICATION NUMBER: JP 2001-272697

PRIOR FILING DATE: 2001-08-02

NUMBER OF SEQ ID NOS: 15109

SEQ ID NO 13774

LENGTH: 157  
TYPE: PRT  
ORGANISM: Streptomyces avermitilis  
US-10-156-761-13774

Query Match  
Best Local Similarity 1.4%; Score 7; DB 15; Length 157;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 410 HOPDGKG 416  
DB 140 HOPDGKG 146

RESULT 6  
US-10-027-450-28  
Sequence 28, Application US/10027450  
Publication No. US20020102715A1  
GENERAL INFORMATION:  
APPLICANT: Falco, Saverio Carl  
APPLICANT: Hite, William D.  
APPLICANT: Kinney, Anthony J.  
APPLICANT: Cahoon, Rebecca E.  
APPLICANT: Rafaleki, J. Antoni  
TITLE OF INVENTION: PLANT BRANCHED CHAIN AMINO ACID BIOSYNTHETIC ENZYMES  
FILE REFERENCE: BB-1126  
CURRENT APPLICATION NUMBER: US/10/027,450  
PRIOR FILING DATE: 2001-12-20  
PRIOR APPLICATION NUMBER: 60/063,423  
PRIOR FILING DATE: 1997 October 28  
NUMBER OF SEQ ID NOS: 54  
SOFTWARE: Microsoft Word Version 7.0A  
SEQ ID NO 28  
LENGTH: 180  
TYPE: PRT  
ORGANISM: Triticum aestivum  
US-10-027-450-28

Query Match  
Best Local Similarity 1.4%; Score 7; DB 14; Length 180;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 349 KVEGEEN 355  
DB 58 KVEGEEN 64

RESULT 7  
US-10-138-787-7  
Sequence 7, Application US/10138787  
Publication No. US20020172984A1  
GENERAL INFORMATION:  
APPLICANT: Holland, Sacha  
APPLICANT: Mbamalu, Geraldine  
APPLICANT: Pawsen, Tony  
TITLE OF INVENTION: OLIGOMERIZED RECEPTORS WHICH AFFECT PATHWAYS REGULATED  
TITLE OF INVENTION: BY TRANSMEMBRANE LIGANDS FOR ELK-RELATED RECEPTOR  
TITLE OF INVENTION: TYROSINE KINASES  
FILE REFERENCE: 11757,23USMO  
CURRENT APPLICATION NUMBER: US/10/138,787  
PRIOR FILING DATE: 2002-05-03  
PRIOR APPLICATION NUMBER: US/09/214,631  
PRIOR FILING DATE: 1999-03-12  
PRIOR APPLICATION NUMBER: PCT/CA97/00473  
PRIOR FILING DATE: 1997-07-04  
PRIOR APPLICATION NUMBER: 60/021,272  
PRIOR FILING DATE: 1996-07-05  
NUMBER OF SEQ ID NOS: 13  
SOFTWARE: Patentln Ver. 2.0  
SEQ ID NO 7  
LENGTH: 233  
TYPE: PRT  
ORGANISM: Homo sapiens

US-10-138-787-7

Query Match  
Best Local Similarity 1.4%; Score 7; DB 14; Length 233;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 148 GPGGAE 154  
DB 74 GPGGAE 80

RESULT 8  
US-09-904-954-2  
Sequence 2, Application US/09904954  
Patent No. US20020010325A1  
GENERAL INFORMATION:  
APPLICANT: BECKMANN, M. P.  
CERRETTI, DOUGLAS P.  
TITLE OF INVENTION: CYTOKINE THAT BINDS THE CELL SURFACE  
RECEPTOR HERK  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: IMMUNEX CORPORATION  
STREET: 51 UNIVERSITY STREET  
CITY: SEATTLE  
STATE: WASHINGTON  
COUNTRY: USA  
ZIP: 98101  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: Apple Macintosh  
OPERATING SYSTEM: Apple System 7.1  
SOFTWARE: Microsoft Word for Apple, Version 5.1a  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/904,954  
FILING DATE: 12-Jul-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/240,124  
FILING DATE: <Unknown>  
APPLICATION NUMBER: US 08/114,426  
FILING DATE: 30-AUG-1993  
APPLICATION NUMBER: US 08/109,745  
FILING DATE: 20-AUG-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: SEESE, KATHRYN A.  
REGISTRATION NUMBER: 32,172  
REFERENCE/DOCKET NUMBER: 2814-C  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 587-0430  
TELEFAX: (206) 233-0644  
TELEX: 756822  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 238 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 2:  
US-09-904-954-2

Query Match  
Best Local Similarity 1.4%; Score 7; DB 9; Length 238;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 148 GPGGAE 154  
DB 78 GPGGAE 84

RESULT 9  
US-10-241-220-72  
Sequence 72, Application US/10241220

```
Publication No. US20030148408A1
GENERAL INFORMATION:
APPLICANT: Frantz, Gretchen
APPLICANT: Hillan, Kenneth J.
APPLICANT: Phillips, Heidi
APPLICANT: Polakis, Paul
APPLICANT: Spencer, Susan
APPLICANT: Williams, P. Mickey
APPLICANT: Wu, Thomas
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND
FILE REFERENCE: P5010R1-US
CURRENT APPLICATION NUMBER: US/10/241,220
CURRENT FILING DATE: 2002-12-13
NUMBER OF SEQ ID NOS: 120
SEQ ID NO 72
LENGTH: 238
TYPE: PRT
ORGANISM: Homo Sapien
US-10-241-220-72
```

```
Query Match      1.4%; Score 7; DB 12; Length 238;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      148 GPGGAAE 154
      |||||
Db      78 GPGGAAE 84
```

```
RESULT 10
US-09-738-626-5155
Sequence 5155, Application US/09738626
Publication No. US20020197605A1
GENERAL INFORMATION:
APPLICANT: NAKAGAWA, SATOSHI
APPLICANT: MIZOGUCHI, HIROSHI
APPLICANT: ANDO, SEIKO
APPLICANT: HAYASHI, MIKIRO
APPLICANT: OCHIAI, KEIKO
APPLICANT: YOKOI, HARUHIKO
APPLICANT: TATEISHI, NAKO
APPLICANT: SENOH, AKIHIRO
APPLICANT: IKEDA, MASATO
APPLICANT: OZAKI, AKIO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-125
CURRENT APPLICATION NUMBER: US/09/738,626
CURRENT FILING DATE: 2000-12-18
PRIOR APPLICATION NUMBER: JP 99/377484
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: JP 00/159162
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: JP 00/280988
PRIOR FILING DATE: 2000-08-03
NUMBER OF SEQ ID NOS: 7059
SOFTWARE: PatentIn ver. 3.0
SEQ ID NO 5155
LENGTH: 278
TYPE: PRT
ORGANISM: Corynebacterium glutamicum
US-09-738-626-5155
```

```
Query Match      1.4%; Score 7; DB 10; Length 278;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      96 RLADGVE 102
      |||||
Db      262 RLADGVE 268
```

```
RESULT 11
US-10-156-761-14202
Sequence 14202, Application US/10156761
Publication No. US20030119018A1
GENERAL INFORMATION:
APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 14202
LENGTH: 300
TYPE: PRT
ORGANISM: Streptomyces avermectilis
US-10-156-761-14202
```

```
Query Match      1.4%; Score 7; DB 15; Length 300;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      96 RLADGVE 102
      |||||
Db      266 RLADGVE 272
```

```
RESULT 12
US-09-967-347-11
Sequence 11, Application US/09967347
Patent No. US20020115829A1
GENERAL INFORMATION:
APPLICANT: Finlay, B. Brett
APPLICANT: Kenny, Brendan
APPLICANT: Stein, Markus
APPLICANT: Domeneberg, Michael S.
APPLICANT: Lai, Li-Ching
TITLE OF INVENTION: PATHOGENIC ESCHERICHIA COLI ASSOCIATED PROTEIN EspA
FILE REFERENCE: 482112.401D1
CURRENT APPLICATION NUMBER: US/09/967,347
CURRENT FILING DATE: 2001-09-28
PRIOR APPLICATION NUMBER: 09/171,517
PRIOR FILING DATE: 1998-10-20
PRIOR APPLICATION NUMBER: PCT/CA97/00265
PRIOR FILING DATE: 1997-04-23
PRIOR APPLICATION NUMBER: 60/015,999
PRIOR FILING DATE: 1996-04-23
NUMBER OF SEQ ID NOS: 15
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 11
LENGTH: 312
TYPE: PRT
ORGANISM: Escherichia coli
US-09-967-347-11
```

```
Query Match      1.4%; Score 7; DB 10; Length 312;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      490 ADDVAKA 496
      |||||
Db      187 ADDVAKA 193
```

RESULT 13  
US-10-092-947A-37  
Sequence 37, Application US/10092947A  
Publication No. US20030134353A1  
GENERAL INFORMATION:  
APPLICANT: WOLFE, Anne M  
APPLICANT: APPEL, Karen F  
APPLICANT: PETERSEN, Jesper F  
APPLICANT: ARNAU, Jose  
APPLICANT: JACOBSEN, Mette D  
TITLE OF INVENTION: MUCOR RECOMBINANT GENE EXPRESSION  
FILE REFERENCE: WOLFE=3  
CURRENT APPLICATION NUMBER: US/10/092,947A  
CURRENT FILING DATE: 2002-12-27  
PRIOR APPLICATION NUMBER: US 60/274,650  
PRIOR FILING DATE: 2001-03-12  
NUMBER OF SEQ ID NOS: 65  
SOFTWARE: PatentIn version 3.2  
SEQ ID NO 37  
LENGTH: 411  
TYPE: PRT  
ORGANISM: Aspergillus niger  
US-10-092-947A-37

Query Match 1.4%; Score 7; DB 12; Length 411;  
Best Local Similarity 100.0%; Pred. No. 2.9e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 436 RAASIVA 442  
Db 368 RAASIVA 374

RESULT 14  
US-09-730-772-14  
Sequence 14, Application US/09730772  
Patent No. US2001001131A1  
GENERAL INFORMATION:  
APPLICANT: Luyten, Frank P.  
APPLICANT: Moos, Jr., Malcolm  
APPLICANT: Chang, Steven Chao-Huan  
TITLE OF INVENTION: CARTILAGE-DERIVED MORPHOGENETIC  
TITLE OF INVENTION: PROTEINS  
NUMBER OF SEQUENCES: 24  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Knobbe, Martens, Olson & Bear  
STREET: 620 Newport Center Drive, 16th Floor  
CITY: Newport Beach  
STATE: CA  
COUNTRY: U.S.A.  
ZIP: 92660  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows  
SOFTWARE: FastSeq for Windows Version 2.0b  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/730,772  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/836,081  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Bartfeld, Neil S  
REGISTRATION NUMBER: 39,901  
REFERENCE/DOCKET NUMBER: NIH099,001APC  
TELEPHONE: 619-235-8550  
TELEFAX: 619-235-0176  
TELEX:  
INFORMATION FOR SEQ ID NO: 14:

SEQUENCE CHARACTERISTICS:  
LENGTH: 436 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-09-730-772-14

Query Match 1.4%; Score 7; DB 9; Length 436;  
Best Local Similarity 100.0%; Pred. No. 3e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 148 GPGGAE 154  
Db 279 GPGGAE 285

RESULT 15  
US-09-735-849-14  
Sequence 14, Application US/09735849  
Patent No. US20010037017A1  
GENERAL INFORMATION:  
APPLICANT: Luyten, Frank P.  
APPLICANT: Moos, Jr., Malcolm  
APPLICANT: Chang, Steven Chao-Huan  
TITLE OF INVENTION: CARTILAGE-DERIVED MORPHOGENETIC  
TITLE OF INVENTION: PROTEINS  
NUMBER OF SEQUENCES: 24  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Knobbe, Martens, Olson & Bear  
STREET: 620 Newport Center Drive, 16th Floor  
CITY: Newport Beach  
STATE: CA  
COUNTRY: U.S.A.  
ZIP: 92660  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows  
SOFTWARE: FastSeq for Windows Version 2.0b  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/735,849  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/836,081  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Bartfeld, Neil S  
REGISTRATION NUMBER: 39,901  
REFERENCE/DOCKET NUMBER: NIH099,001APC  
TELEPHONE: 619-235-8550  
TELEFAX: 619-235-0176  
TELEX:  
INFORMATION FOR 'SEQ ID NO: 14':  
SEQUENCE CHARACTERISTICS:  
LENGTH: 436 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-09-735-849-14

Query Match 1.4%; Score 7; DB 9; Length 436;  
Best Local Similarity 100.0%; Pred. No. 3e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 148 GPGGAE 154  
Db 279 GPGGAE 285

RESULT 16  
US-10-062-848-2  
; Sequence 2, Application US/10062848  
; Publication No. US20030092155A1  
; GENERAL INFORMATION:  
; APPLICANT: KOSTREMA, Dirk  
; APPLICANT: PASAMONTES, Luis  
; APPLICANT: TOMSCHY, Andrea  
; APPLICANT: van LOON, Adolphus  
; APPLICANT: VOGEL, Kurt  
; APPLICANT: WYSS, Markus  
; TITLE OF INVENTION: MODIFIED PHYTASES  
; FILE REFERENCE: Modified Phytases  
; CURRENT APPLICATION NUMBER: US/10/062,848  
; PRIOR FILING DATE: 2002-02-01  
; PRIOR APPLICATION NUMBER: 09/044,718  
; PRIOR FILING DATE: 1998-03-19  
; PRIOR APPLICATION NUMBER: EP 97810175.6  
; PRIOR FILING DATE: 1997-03-25  
; NUMBER OF SEQ ID NOS: 82  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 438  
; TYPE: PRT  
; ORGANISM: Aspergillus terreus  
US-10-062-848-2

Query Match 1.4%; Score 7; DB 15; Length 438;  
Best Local Similarity 100.0%; Pred. No. 3e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 487 QLSADVV 493  
DB 223 QLSADVV 229

RESULT 17  
US-09-925-637-20  
; Sequence 20, Application US/09925637  
; Patent No. US2002010338A1  
; GENERAL INFORMATION:  
; APPLICANT: Choi  
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Polypeptides  
; FILE REFERENCE: PB560  
; CURRENT APPLICATION NUMBER: US/09/925,637  
; PRIOR FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: PCT/US00/23773  
; PRIOR FILING DATE: 2000-08-31  
; PRIOR APPLICATION NUMBER: US 60/151,933  
; PRIOR FILING DATE: 1999-09-01  
; PRIOR APPLICATION NUMBER: US 08/781,986  
; PRIOR FILING DATE: 1997-01-03  
; PRIOR APPLICATION NUMBER: US 08/956,171  
; PRIOR FILING DATE: 1997-10-20  
; PRIOR APPLICATION NUMBER: US 60/009,861  
; PRIOR FILING DATE: 1996-01-06  
; NUMBER OF SEQ ID NOS: 74  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 20  
; LENGTH: 439  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-925-637-20

Query Match 1.4%; Score 7; DB 10; Length 439;  
Best Local Similarity 100.0%; Pred. No. 3e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 436 RAASIVA 442  
DB 301 RAASIVA 307

RESULT 18  
US-10-084-205-20  
; Sequence 20, Application US/10084205  
; Publication No. US20030049648A1  
; GENERAL INFORMATION:  
; APPLICANT: Choi, Gil  
; TITLE OF INVENTION: 37 Staphylococcus aureus Genes and Polypeptides  
; FILE REFERENCE: PB515P1  
; CURRENT APPLICATION NUMBER: US/10/084,205  
; PRIOR FILING DATE: 2002-02-28  
; PRIOR APPLICATION NUMBER: PCT/US00/23773  
; PRIOR FILING DATE: 2000-08-31  
; PRIOR APPLICATION NUMBER: 60/151,933  
; PRIOR FILING DATE: 1999-09-01  
; NUMBER OF SEQ ID NOS: 74  
; SOFTWARE: PatentIn Ver. 3.1  
; SEQ ID NO 20  
; LENGTH: 439  
; TYPE: PRT  
; ORGANISM: Staphylococcus aureus  
US-10-084-205-20

Query Match 1.4%; Score 7; DB 15; Length 439;  
Best Local Similarity 100.0%; Pred. No. 3e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 436 RAASIVA 442  
DB 301 RAASIVA 307

RESULT 19  
US-10-062-848-15  
; Sequence 15, Application US/10062848  
; Publication No. US20030092155A1  
; GENERAL INFORMATION:  
; APPLICANT: KOSTREMA, Dirk  
; APPLICANT: PASAMONTES, Luis  
; APPLICANT: TOMSCHY, Andrea  
; APPLICANT: van LOON, Adolphus  
; APPLICANT: VOGEL, Kurt  
; APPLICANT: WYSS, Markus  
; TITLE OF INVENTION: MODIFIED PHYTASES  
; FILE REFERENCE: Modified Phytases  
; CURRENT APPLICATION NUMBER: US/10/062,848  
; PRIOR FILING DATE: 2002-02-01  
; PRIOR APPLICATION NUMBER: 09/044,718  
; PRIOR FILING DATE: 1998-03-19  
; PRIOR APPLICATION NUMBER: EP 97810175.6  
; PRIOR FILING DATE: 1997-03-25  
; NUMBER OF SEQ ID NOS: 82  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 15  
; LENGTH: 450  
; TYPE: PRT  
; ORGANISM: Aspergillus terreus  
US-10-062-848-15

Query Match 1.4%; Score 7; DB 15; Length 450;  
Best Local Similarity 100.0%; Pred. No. 3.1e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 487 QLSADVV 493  
DB 235 QLSADVV 241

RESULT 20  
US-09-843-250-24  
; Sequence 24, Application US/09843250  
; Publication No. US20030022335A1  
; GENERAL INFORMATION:  
; APPLICANT: Parales, R.

APPLICANT: Gibson, D.  
APPLICANT: Resnick, S.  
APPLICANT: Lee, K.  
TITLE OF INVENTION: No. US20030022335A1el naphthalene dioxygenase and methods for the  
FILE REFERENCE: 875.006US2  
CURRENT APPLICATION NUMBER: US/09/843,250  
CURRENT FILING DATE: 2001-04-26  
PRIOR APPLICATION NUMBER: PCT/US99/25079  
PRIOR FILING DATE: 1999-10-26  
PRIOR APPLICATION NUMBER: US 60/105,575  
PRIOR FILING DATE: 1998-10-26  
NUMBER OF SEQ ID NOS: 65  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 24  
LENGTH: 451  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: A polypeptide encoded by SEQ ID NO:13.  
US-09-843-250-24

Query Match 1.4%; Score 7; DB 11; Length 451;  
Best Local Similarity 100.0%; Pred. No. 3.1e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 81 PAKVVK 87  
| | | | |  
Db 189 PAKVVK 195

RESULT 21  
US-10-278-481-18  
Sequence 18, Application US/10278481  
Publication No. US20030113803A1  
GENERAL INFORMATION:  
APPLICANT: KARO BIO AB  
TITLE OF INVENTION: Opihan receptor  
NUMBER OF SEQUENCES: 19  
STREET: c/o Center for Biotechnology and Department  
of Medical Nutrition, Karolinska Nutrition, Karolinska  
Institute  
CITY: Huddinge  
COUNTRY: Sweden  
ZIP: S-14186  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/278,481  
FILING DATE: 23-Oct-2002  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/09/333,057  
FILING DATE: <Unknown>  
APPLICATION NUMBER: 08/836,620  
FILING DATE: <Unknown>  
APPLICATION NUMBER: GB 9518272.1  
FILING DATE: 08-SEP-1995  
APPLICATION NUMBER: GB 9605550.4  
FILING DATE: 15-MAR-1996  
APPLICATION NUMBER: GB 9607532.0  
FILING DATE: 11-APR-1996  
APPLICATION NUMBER: GB 9609576.5  
FILING DATE: 08-MAY-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: GUSTAFSSON, Jan-Ake  
INFORMATION FOR SEQ ID NO: 18:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 518 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
ORIGINAL SOURCE:

ORGANISM: Homo sapiens  
SEQUENCE DESCRIPTION: SEQ ID NO: 18:  
US-10-278-481-18

Query Match 1.4%; Score 7; DB 15; Length 518;  
Best Local Similarity 100.0%; Pred. No. 3.5e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 148 GPGGAE 154  
| | | | |  
Db 466 GPGGAE 472

RESULT 22  
US-10-211-239-1  
Sequence 1, Application US/10211239  
Publication No. US20030103965A1  
GENERAL INFORMATION:  
APPLICANT: Jung, Birgit  
APPLICANT: Kraut, No. US20030103965A1bert  
APPLICANT: Mueller, Stefan  
TITLE OF INVENTION: Method for identifying Substances Which Positively  
Influence Inflammatory Conditions  
FILE REFERENCE: 0652.2340001  
CURRENT APPLICATION NUMBER: US/10/211,239  
CURRENT FILING DATE: 2002-08-05  
PRIOR APPLICATION NUMBER: US 60/315,775  
PRIOR FILING DATE: 2001-08-30  
PRIOR APPLICATION NUMBER: EP 01119003.0  
PRIOR FILING DATE: 2001-08-06  
NUMBER OF SEQ ID NOS: 10  
SOFTWARE: Patentin version 3.1  
SEQ ID NO 1  
LENGTH: 519  
TYPE: PRT  
ORGANISM: Homo Sapiens  
US-10-211-239-1

Query Match 1.4%; Score 7; DB 15; Length 519;  
Best Local Similarity 100.0%; Pred. No. 3.6e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 148 GPGGAE 154  
| | | | |  
Db 464 GPGGAE 470

RESULT 23  
US-10-156-761-14029  
Sequence 14029, Application US/10156761  
Publication No. US20030119018A1  
GENERAL INFORMATION:  
APPLICANT: OKURA, SATOSHI  
APPLICANT: IKEDA, HARUO  
APPLICANT: ISHIRAWA, JUN  
APPLICANT: HORIKAWA, HIROSHI  
APPLICANT: SHIBA, TADAYOSHI  
APPLICANT: SAKAKI, YOSHIYUKI  
APPLICANT: HATTORI, MASAHIRA  
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
FILE REFERENCE: 249-262  
CURRENT APPLICATION NUMBER: US/10/156,761  
CURRENT FILING DATE: 2002-05-29  
PRIOR APPLICATION NUMBER: JP 2001-204089  
PRIOR FILING DATE: 2001-05-30  
PRIOR APPLICATION NUMBER: JP 2001-272697  
PRIOR FILING DATE: 2001-08-02  
NUMBER OF SEQ ID NOS: 15109  
SEQ ID NO 14029  
LENGTH: 549  
TYPE: PRT  
ORGANISM: Streptomyces avermitilis  
US-10-156-761-14029



Query Match 1.4%; Score 7; DB 15; Length 549;  
Best Local Similarity 100.0%; Pred. No. 3.7e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 94 VMRLADG 100  
Db 326 VMRLADG 332

## RESULT 24

US-09-815-242-12111  
; Sequence 12111, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Karl L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; PRIOR FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12111  
; LENGTH: 611  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-815-242-12111

Query Match 1.4%; Score 7; DB 9; Length 611;  
Best Local Similarity 100.0%; Pred. No. 4.1e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 96 RLADGVE 102  
Db 427 RLADGVE 433

RESULT 25  
US-10-156-761-9578  
; Sequence 9578, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262

; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 9578  
; LENGTH: 724  
; TYPE: PRT  
; ORGANISM: Streptomyces avermitilis  
US-10-156-761-9578

Query Match 1.4%; Score 7; DB 15; Length 724;  
Best Local Similarity 100.0%; Pred. No. 4.8e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 97 LADGVEY 103  
Db 260 LADGVEY 266

RESULT 26  
US-09-995-587A-11  
; Sequence 11, Application US/0995587A  
; Patent No. US20020127681A1  
; GENERAL INFORMATION:  
; APPLICANT: VAN HIJUM, SACHA ADRIANUS FOKKE TACO  
; APPLICANT: VAN GEEL-SCHUTTEN, GERTRIDINA HENDRIKA  
; APPLICANT: DIJKHUIZEN, LUBBERT  
; APPLICANT: RAHAOUI, HAKIM  
; TITLE OF INVENTION: NOVEL FRUCTOSYLTRANSFERASES  
; FILE REFERENCE: B043667-CIP  
; CURRENT APPLICATION NUMBER: US/09/995,587A  
; CURRENT FILING DATE: 2002-02-28  
; PRIOR APPLICATION NUMBER: 09/604,958  
; PRIOR FILING DATE: 2000-06-28  
; PRIOR APPLICATION NUMBER: EPO 00201872.9  
; PRIOR FILING DATE: 2000-05-25  
; NUMBER OF SEQ ID NOS: 40  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 11  
; LENGTH: 792  
; TYPE: PRT  
; ORGANISM: Lactobacillus reuteri  
; NAME/KEY: MOD RES  
; LOCATION: (495)..(496)  
; OTHER INFORMATION: Any amino acid  
; FEATURE:  
; NAME/KEY: MOD RES  
; LOCATION: (737)  
; OTHER INFORMATION: Thr or Pro  
US-09-995-587A-11

Query Match 1.4%; Score 7; DB 10; Length 792;  
Best Local Similarity 100.0%; Pred. No. 5.2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 343 GALGILK 349  
Db 461 GALGILK 467

RESULT 27  
US-09-882-227-308  
; Sequence 308, Application US/09882227  
; Publication No. US20030158396A1  
; GENERAL INFORMATION:  
; APPLICANT: Kleenhouw, Harold  
; APPLICANT: Al-Garawi, Amal  
; APPLICANT: Miller, Charles  
; APPLICANT: Tomb, Jean-Francois

APPLICANT: Oomen, Raymond P.  
TITLE OF INVENTION: Identification of Polynucleotides  
TITLE OF INVENTION: Encoding No. US2003015896A1 Helicobacter Polypeptides in the  
TITLE OF INVENTION: Genome  
FILE REFERENCE: 06132/047002  
CURRENT APPLICATION NUMBER: US/09/882,227  
CURRENT FILING DATE: 2001-06-15  
PRIOR APPLICATION NUMBER: US 08/902,615  
PRIOR FILING DATE: 1997-07-29  
NUMBER OF SEQ ID NOS: 638  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 308  
LENGTH: 844  
TYPE: PRT  
ORGANISM: Helicobacter pylori  
US-09-882-227-308

Query Match 1.4%; Score 7; DB 12; Length 844;  
Best Local Similarity 100.0%; Pred. No. 5.5e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 357 EYSHKQ 363  
Db 553 EYSHKQ 559

RESULT 28  
US-08-945-749-1  
Sequence 1, Application US/08945749  
Publication No. US20020138880A1  
GENERAL INFORMATION:  
APPLICANT: GARDNER, Richard C  
APPLICANT: MacDIAMID, Colin W  
APPLICANT: HAY, Robert J  
APPLICANT: Auckland Uniservices Limited  
APPLICANT: New Zealand Pastoral Agriculture Research Institut  
TITLE OF INVENTION: ALUMINUM RESISTANCE GENE  
FILE REFERENCE: 08/945,749  
CURRENT APPLICATION NUMBER: US/08/945,749  
CURRENT FILING DATE: 1998-01-12  
EARLIER APPLICATION NUMBER: PCT/NZ96/0035  
EARLIER FILING DATE: 1996-05-01  
EARLIER APPLICATION NUMBER: NZ 272039  
EARLIER FILING DATE: 1995-05-01  
NUMBER OF SEQ ID NOS: 16  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 1  
LENGTH: 859  
TYPE: PRT  
ORGANISM: Saccharomyces cerevisiae  
US-08-945-749-1

Query Match 1.4%; Score 7; DB 8; Length 859;  
Best Local Similarity 100.0%; Pred. No. 5.6e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 349 KVEEEN 355  
Db 388 KVEEEN 394

RESULT 29  
US-08-781-986A-5235  
Sequence 5235, Application US/08781986A  
Publication No. US2003005436A1  
GENERAL INFORMATION:  
APPLICANT: Charles Kunsch  
TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences  
NUMBER OF SEQUENCES: 5255  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville

STATE: Maryland  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/781,986A  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Benson, Bob  
REGISTRATION NUMBER: 30,446  
REFERENCE/DOCKET NUMBER: PB248PP  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512  
INFORMATION FOR SEQ ID NO: 5235:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 886 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-781-986A-5235

Query Match 1.4%; Score 7; DB 8; Length 886;  
Best Local Similarity 100.0%; Pred. No. 5.8e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 31 AOPKST 37  
Db 295 AOPKST 301

RESULT 30  
US-10-188-308-21  
Sequence 21, Application US/10188308  
Publication No. US20030077735A1  
GENERAL INFORMATION:  
APPLICANT: PARDO-FERNANDEZ, LUIS ANGEL  
APPLICANT: STUMMER, WALTER  
APPLICANT: BECKH, SYMNOVE  
APPLICANT: BRUGEMANN, ANDREA  
APPLICANT: FERNANDEZ-MIRANDA, DONATO DEL CAMINO  
APPLICANT: PEREZ, ARACELI SANCHEZ  
APPLICANT: WESSELHO, RUDIGER  
TITLE OF INVENTION: NOVEL HUMAN K+ ION CHANNEL AND THERAPEUTIC  
TITLE OF INVENTION: APPLICATIONS  
TITLE OF INVENTION: THEROP  
FILE REFERENCE: MPG-8  
CURRENT APPLICATION NUMBER: US/10/188,308  
CURRENT FILING DATE: 2002-07-01  
PRIOR APPLICATION NUMBER: US/09/694,777  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: PCT/EP99/02695  
PRIOR FILING DATE: 1999-04-21  
PRIOR APPLICATION NUMBER: EP 98 10 7268.9  
PRIOR FILING DATE: 1998-04-21  
NUMBER OF SEQ ID NOS: 24  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 21  
LENGTH: 960  
TYPE: PRT  
ORGANISM: Bovine sp.  
US-10-188-308-21

Query Match 1.4%; Score 7; DB 15; Length 960;

Best Local Similarity 100.0%; Pred. No. 6.2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 380 QCAPKTP 386  
|||||  
DB 200 QCAPKTP 206

RESULT 31  
US-10-188-296-21

; Sequence 21, Application US/10188296  
; Publication No. US20030087377A1

; GENERAL INFORMATION:

; APPLICANT: PARDO-FERNANDEZ, LUIS ANGEL

; APPLICANT: STUMMER, WALTER

; APPLICANT: BECKH, SYNOVE

; APPLICANT: BRUGEMANN, ANDREA

; APPLICANT: FERNANDEZ-MIRANDA, DONATO DEL CAMINO

; APPLICANT: PEREZ, ARACELI SANCHEZ

; APPLICANT: WESELOH, RUDIGER

; TITLE OF INVENTION: NOVEL HUMAN K+ ION CHANNEL AND THERAPEUTIC APPLICATIONS

; FILE REFERENCE: MPG-8

; CURRENT APPLICATION NUMBER: US/10/188,296

; CURRENT FILING DATE: 2002-07-01

; PRIOR APPLICATION NUMBER: US/09/694,777

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: PCT/EP99/02695

; PRIOR FILING DATE: 1999-04-21

; PRIOR APPLICATION NUMBER: EP 98 10 7268.9

; PRIOR FILING DATE: 1998-04-21

; NUMBER OF SEQ ID NOS: 24

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 21

; LENGTH: 960

; TYPE: PRT

; ORGANISM: Bovine sp.

US-10-188-296-21

Query Match 1.4%; Score 7; DB 15; Length 960;

Best Local Similarity 100.0%; Pred. No. 6.2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 380 QCAPKTP 386  
|||||  
DB 200 QCAPKTP 206

RESULT 32  
US-10-188-341-21

; Sequence 21, Application US/10188341

; Publication No. US20030087378A1

; GENERAL INFORMATION:

; APPLICANT: PARDO-FERNANDEZ, LUIS ANGEL

; APPLICANT: STUMMER, WALTER

; APPLICANT: BECKH, SYNOVE

; APPLICANT: BRUGEMANN, ANDREA

; APPLICANT: FERNANDEZ-MIRANDA, DONATO DEL CAMINO

; APPLICANT: PEREZ, ARACELI SANCHEZ

; APPLICANT: WESELOH, RUDIGER

; TITLE OF INVENTION: NOVEL HUMAN K+ ION CHANNEL AND THERAPEUTIC APPLICATIONS

; FILE REFERENCE: MPG-8

; CURRENT APPLICATION NUMBER: US/10/188,341

; CURRENT FILING DATE: 2002-07-01

; PRIOR APPLICATION NUMBER: US/09/694,777

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: PCT/EP99/02695

; PRIOR FILING DATE: 1999-04-21

; PRIOR APPLICATION NUMBER: EP 98 10 7268.9

; PRIOR FILING DATE: 1998-04-21

; NUMBER OF SEQ ID NOS: 24

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 21  
; LENGTH: 960  
; TYPE: PRT  
; ORGANISM: Bovine sp.  
US-10-188-341-21

Query Match 1.4%; Score 7; DB 15; Length 960;  
Best Local Similarity 100.0%; Pred. No. 6.2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 380 QCAPKTP 386  
|||||  
DB 200 QCAPKTP 206

RESULT 33  
US-10-188-297-21

; Sequence 21, Application US/10188297

; Publication No. US20030092120A1

; GENERAL INFORMATION:

; APPLICANT: PARDO-FERNANDEZ, LUIS ANGEL

; APPLICANT: STUMMER, WALTER

; APPLICANT: BECKH, SYNOVE

; APPLICANT: BRUGEMANN, ANDREA

; APPLICANT: FERNANDEZ-MIRANDA, DONATO DEL CAMINO

; APPLICANT: PEREZ, ARACELI SANCHEZ

; APPLICANT: WESELOH, RUDIGER

; TITLE OF INVENTION: NOVEL HUMAN K+ ION CHANNEL AND THERAPEUTIC APPLICATIONS

; FILE REFERENCE: MPG-8

; CURRENT APPLICATION NUMBER: US/10/188,297

; CURRENT FILING DATE: 2002-07-01

; PRIOR APPLICATION NUMBER: US/09/694,777

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: PCT/EP99/02695

; PRIOR FILING DATE: 1999-04-21

; PRIOR APPLICATION NUMBER: EP 98 10 7268.9

; PRIOR FILING DATE: 1998-04-21

; NUMBER OF SEQ ID NOS: 24

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 21

; LENGTH: 960

; TYPE: PRT

; ORGANISM: Bovine sp.

US-10-188-297-21

Query Match 1.4%; Score 7; DB 15; Length 960;

Best Local Similarity 100.0%; Pred. No. 6.2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 380 QCAPKTP 386  
|||||  
DB 200 QCAPKTP 206

RESULT 34  
US-10-188-308-3

; Sequence 3, Application US/10188308

; Publication No. US2003007735A1

; GENERAL INFORMATION:

; APPLICANT: PARDO-FERNANDEZ, LUIS ANGEL

; APPLICANT: STUMMER, WALTER

; APPLICANT: BECKH, SYNOVE

; APPLICANT: BRUGEMANN, ANDREA

; APPLICANT: FERNANDEZ-MIRANDA, DONATO DEL CAMINO

; APPLICANT: PEREZ, ARACELI SANCHEZ

; APPLICANT: WESELOH, RUDIGER

; TITLE OF INVENTION: NOVEL HUMAN K+ ION CHANNEL AND THERAPEUTIC

; FILE REFERENCE: MPG-8

; CURRENT APPLICATION NUMBER: US/10/188,308

; CURRENT FILING DATE: 2002-07-01

PRIOR APPLICATION NUMBER: US/09/694,777  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: PCT/EP99/02695  
PRIOR FILING DATE: 1999-04-21  
PRIOR APPLICATION NUMBER: EP 98 10 7268.9  
PRIOR FILING DATE: 1998-04-21  
NUMBER OF SEQ ID NOS: 24  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 3  
LENGTH: 962  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-188-308-3

Query Match 1.4% Score 7; DB 15; Length 962;  
Best Local Similarity 100.0%; Pred. No. 6.2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 380 QEAPKTP 386  
|||||  
Db 200 QEAPKTP 206

RESULT 35  
US-10-188-308-24  
Sequence 24, Application US/10188308  
Publication No. US20030077735A1  
GENERAL INFORMATION:  
APPLICANT: PARDO-FERNANDEZ, LUIS ANGEL  
APPLICANT: STUMMER, WALTER  
APPLICANT: BECKH, SYNNOVE  
APPLICANT: BRUGEMANN, ANDREA  
APPLICANT: FERNANDEZ-MIRANDA, DONATO DEL CAMINO  
APPLICANT: PEREZ, ARACELI SANCHEZ  
APPLICANT: WESELOH, RUDIGER  
TITLE OF INVENTION: NOVEL HUMAN K+ ION CHANNEL AND THERAPEUTIC  
TITLE OF INVENTION: APPLICATIONS  
FILE REFERENCE: THEREOF  
FILE REFERENCE: MPG-8  
CURRENT APPLICATION NUMBER: US/10/188,308  
CURRENT FILING DATE: 2002-07-01  
PRIOR APPLICATION NUMBER: US/09/694,777  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: PCT/EP99/02695  
PRIOR FILING DATE: 1999-04-21  
PRIOR APPLICATION NUMBER: EP 98 10 7268.9  
PRIOR FILING DATE: 1998-04-21  
NUMBER OF SEQ ID NOS: 24  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 24  
LENGTH: 962  
TYPE: PRT  
ORGANISM: Rattus sp.  
US-10-188-308-24

Query Match 1.4% Score 7; DB 15; Length 962;  
Best Local Similarity 100.0%; Pred. No. 6.2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 380 QEAPKTP 386  
|||||  
Db 200 QEAPKTP 206

RESULT 36  
US-10-188-296-3  
Sequence 3, Application US/10188296  
Publication No. US20030087377A1  
GENERAL INFORMATION:  
APPLICANT: PARDO-FERNANDEZ, LUIS ANGEL  
APPLICANT: STUMMER, WALTER  
APPLICANT: BECKH, SYNNOVE  
APPLICANT: BRUGEMANN, ANDREA

APPLICANT: FERNANDEZ-MIRANDA, DONATO DEL CAMINO  
APPLICANT: PEREZ, ARACELI SANCHEZ  
APPLICANT: WESELOH, RUDIGER  
TITLE OF INVENTION: NOVEL HUMAN K+ ION CHANNEL AND THERAPEUTIC APPLICATIONS  
TITLE OF INVENTION: THEREOF  
FILE REFERENCE: MPG-8  
CURRENT APPLICATION NUMBER: US/10/188,296  
CURRENT FILING DATE: 2002-07-01  
PRIOR APPLICATION NUMBER: US/09/694,777  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: PCT/EP99/02695  
PRIOR FILING DATE: 1999-04-21  
PRIOR APPLICATION NUMBER: EP 98 10 7268.9  
PRIOR FILING DATE: 1998-04-21  
NUMBER OF SEQ ID NOS: 24  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 3  
LENGTH: 962  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-188-296-3

Query Match 1.4% Score 7; DB 15; Length 962;  
Best Local Similarity 100.0%; Pred. No. 6.2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 380 QEAPKTP 386  
|||||  
Db 200 QEAPKTP 206

RESULT 37  
US-10-188-296-24  
Sequence 24, Application US/10188296  
Publication No. US20030087377A1  
GENERAL INFORMATION:  
APPLICANT: PARDO-FERNANDEZ, LUIS ANGEL  
APPLICANT: STUMMER, WALTER  
APPLICANT: BECKH, SYNNOVE  
APPLICANT: BRUGEMANN, ANDREA  
APPLICANT: FERNANDEZ-MIRANDA, DONATO DEL CAMINO  
APPLICANT: PEREZ, ARACELI SANCHEZ  
APPLICANT: WESELOH, RUDIGER  
TITLE OF INVENTION: NOVEL HUMAN K+ ION CHANNEL AND THERAPEUTIC APPLICATIONS  
TITLE OF INVENTION: THEREOF  
FILE REFERENCE: MPG-8  
CURRENT APPLICATION NUMBER: US/10/188,296  
CURRENT FILING DATE: 2002-07-01  
PRIOR APPLICATION NUMBER: US/09/694,777  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: PCT/EP99/02695  
PRIOR FILING DATE: 1999-04-21  
PRIOR APPLICATION NUMBER: EP 98 10 7268.9  
PRIOR FILING DATE: 1998-04-21  
NUMBER OF SEQ ID NOS: 24  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 24  
LENGTH: 962  
TYPE: PRT  
ORGANISM: Rattus sp.  
US-10-188-296-24

Query Match 1.4% Score 7; DB 15; Length 962;  
Best Local Similarity 100.0%; Pred. No. 6.2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 380 QEAPKTP 386  
|||||  
Db 200 QEAPKTP 206

RESULT 38  
US-10-188-341-3

```
; Sequence 3, Application US/10188341
; Publication No. US20030087378A1
; GENERAL INFORMATION:
; APPLICANT: PARDO-FERNANDEZ, LUIS ANGEL
; APPLICANT: STUHRER, WALTER
; APPLICANT: BECK, SYMNOVE
; APPLICANT: BRUGEMANN, ANDREA
; APPLICANT: FERNANDEZ-MIRANDA, DONATO DEL CAMINO
; APPLICANT: PEREZ, ARACELI SANCHEZ
; APPLICANT: WESELOH, RUDIGER
; TITLE OF INVENTION: NOVEL HUMAN K+ ION CHANNEL AND THERAPEUTIC APPLICATIONS
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: MPG-8
; CURRENT APPLICATION NUMBER: US/10/188,341
; CURRENT FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: US/09/694,777
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: PCT/EP99/02695
; PRIOR FILING DATE: 1999-04-21
; PRIOR APPLICATION NUMBER: EP 98 10 7268.9
; PRIOR FILING DATE: 1998-04-21
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 962
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-188-341-3
```

```
Query Match
Best Local Similarity 1.4%; Score 7; DB 15; Length 962;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 380 OEAPKTP 386
Db 200 OEAPKTP 206
```

```
RESULT 39
US-10-188-341-24
; Sequence 24, Application US/10188341
; Publication No. US20030087378A1
; GENERAL INFORMATION:
; APPLICANT: PARDO-FERNANDEZ, LUIS ANGEL
; APPLICANT: STUHRER, WALTER
; APPLICANT: BECK, SYMNOVE
; APPLICANT: BRUGEMANN, ANDREA
; APPLICANT: FERNANDEZ-MIRANDA, DONATO DEL CAMINO
; APPLICANT: PEREZ, ARACELI SANCHEZ
; APPLICANT: WESELOH, RUDIGER
; TITLE OF INVENTION: NOVEL HUMAN K+ ION CHANNEL AND THERAPEUTIC APPLICATIONS
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: MPG-8
; CURRENT APPLICATION NUMBER: US/10/188,341
; CURRENT FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: US/09/694,777
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: PCT/EP99/02695
; PRIOR FILING DATE: 1999-04-21
; PRIOR APPLICATION NUMBER: EP 98 10 7268.9
; PRIOR FILING DATE: 1998-04-21
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 24
; LENGTH: 962
; TYPE: PRT
; ORGANISM: Rattus sp.
US-10-188-341-24
```

```
Query Match
Best Local Similarity 1.4%; Score 7; DB 15; Length 962;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 380 OEAPKTP 386
Db 200 OEAPKTP 206
```

```
RESULT 40
US-10-188-297-3
; Sequence 3, Application US/10188297
; Publication No. US20030092120A1
; GENERAL INFORMATION:
; APPLICANT: PARDO-FERNANDEZ, LUIS ANGEL
; APPLICANT: STUHRER, WALTER
; APPLICANT: BECK, SYMNOVE
; APPLICANT: BRUGEMANN, ANDREA
; APPLICANT: FERNANDEZ-MIRANDA, DONATO DEL CAMINO
; APPLICANT: PEREZ, ARACELI SANCHEZ
; APPLICANT: WESELOH, RUDIGER
; TITLE OF INVENTION: NOVEL HUMAN K+ ION CHANNEL AND THERAPEUTIC APPLICATIONS
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: MPG-8
; CURRENT APPLICATION NUMBER: US/10/188,297
; CURRENT FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: US/09/694,777
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: PCT/EP99/02695
; PRIOR FILING DATE: 1999-04-21
; PRIOR APPLICATION NUMBER: EP 98 10 7268.9
; PRIOR FILING DATE: 1998-04-21
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 962
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-188-297-3
```

```
Query Match
Best Local Similarity 1.4%; Score 7; DB 15; Length 962;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 380 OEAPKTP 386
Db 200 OEAPKTP 206
```

```
Search completed: August 27, 2003, 18:54:16
Job time : 72 secs
```

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Comugen Ltd.

## OM protein - protein search, using sw model

Run on: August 27, 2003, 18:32:47 ; Search time 86 Seconds

(without alignments)  
926.520 Million cell updates/sec

Title: US-10-088-045-2

Perfect score: 502  
Sequence: 1 MSKPLTIKTTLLICALSALML.....NKGQISADVAKAKKTKEP 502Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 1107863 seqs, 158726573 residues

Word size : 0  
Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database : A.GeneSeq\_19Jun03:\*

1: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1980.DAT:\*  
2: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1981.DAT:\*  
3: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1982.DAT:\*  
4: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1983.DAT:\*  
5: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1984.DAT:\*  
6: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1985.DAT:\*  
7: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1986.DAT:\*  
8: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1987.DAT:\*  
9: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1988.DAT:\*  
10: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1989.DAT:\*  
11: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1990.DAT:\*  
12: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1991.DAT:\*  
13: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1992.DAT:\*  
14: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1993.DAT:\*  
15: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1994.DAT:\*  
16: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1995.DAT:\*  
17: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1996.DAT:\*  
18: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1997.DAT:\*  
19: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1998.DAT:\*  
20: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1999.DAT:\*  
21: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA2000.DAT:\*  
22: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA2001.DAT:\*  
23: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA2002.DAT:\*  
24: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA2003.DAT:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	502	100.0	502	22	AAB67668	Amino acid sequenc
2	502	100.0	502	22	AAB67669	Amino acid sequenc
3	197	39.2	197	23	AA017561	M. catarrhalis MCA1
4	34	6.8	392	24	AB077627	N. gonorrhoeae ami
5	8	1.6	10	20	AA05975	Human cancer anti
6	8	1.6	415	22	ABG18127	Novel human diagno
7	8	1.6	497	21	AA174923	Neisseria gonorth
8	8	1.6	497	21	AA174924	Neisseria meningit
9	8	1.6	497	21	AA174925	Neisseria meningit

10	8	1.6	508	24	ABP80199	N. gonorrhoeae ami
11	8	1.6	697	21	AA193290	Amino acid sequenc
12	8	1.6	697	21	AA193303	Amino acid sequenc
13	8	1.6	995	23	ABP65660	Bifidobacterium 10
14	7	1.4	33	21	AA132246	Partial sequence o
15	7	1.4	39	22	AA193319	Human protein HP10
16	7	1.4	71	22	ABG55536	Human liver peptid
17	7	1.4	71	22	ABG40293	Peptide #7799 enco
18	7	1.4	71	22	AA161086	Human brain expres
19	7	1.4	71	22	AA173789	Human bone marrow
20	7	1.4	71	22	AA133976	Peptide #803 enco
21	7	1.4	71	22	ABG43675	Human peptide enco
22	7	1.4	90	22	ABP60712	Drosophila melanog
23	7	1.4	94	22	AA010927	Human polypeptide
24	7	1.4	104	17	AA107245	HIV-1 group O stra
25	7	1.4	105	22	ABG26287	Novel human diagno
26	7	1.4	111	16	AA172394	Agmenellum quadrip
27	7	1.4	114	23	ABP89496	Human polypeptide
28	7	1.4	118	21	AA127869	Sequence homologou
29	7	1.4	124	22	ABG08519	Novel human diagno
30	7	1.4	124	22	AA001520	Human polypeptide
31	7	1.4	130	23	ABP55608	Lactococcus lactis
32	7	1.4	144	21	AA107374	Arabidopsis thalia
33	7	1.4	149	21	AA107373	Arabidopsis thalia
34	7	1.4	156	21	AA125190	Eucalyptus grandis
35	7	1.4	167	22	ABP58068	Drosophila melanog
36	7	1.4	172	23	ABP38126	Staphylococcus epi
37	7	1.4	180	21	AA128428	Wheat branched cha
38	7	1.4	180	22	AA186749	Human immune/haema
39	7	1.4	185	22	ABP71959	Drosophila melanog
40	7	1.4	187	22	ABP71208	Drosophila melanog
41	7	1.4	189	22	AA162704	Propionibacterium
42	7	1.4	224	19	AA166164	Myobacterium tube
43	7	1.4	228	22	ABP62268	Drosophila melanog
44	7	1.4	228	22	ABG00053	Novel human diagno
45	7	1.4	234	16	AA182605	Eph transmembrane
46	7	1.4	234	24	AA182605	Human secreted pro
47	7	1.4	238	24	ABP00092	Human gene 82 enco
48	7	1.4	238	16	AA174481	Human hec-L protei
49	7	1.4	238	23	AA179439	Braesc/colon cance
50	7	1.4	245	23	ABP55468	Lactococcus lactis
51	7	1.4	250	22	ABP69076	Drosophila melanog
52	7	1.4	259	21	AA124855	Arabidopsis thalia
53	7	1.4	276	18	AA120822	H. pylori secreted
54	7	1.4	278	22	AA191401	C glutamicum prote
55	7	1.4	278	22	AA195498	Human protein sequ
56	7	1.4	278	22	AA179157	Corynebacterium gl
57	7	1.4	281	22	ABP61929	Novel human diagno
58	7	1.4	287	22	ABG28799	Novel human diagno
59	7	1.4	288	22	AA164161	Propionibacterium
60	7	1.4	293	23	ABP53671	Lactococcus lactis
61	7	1.4	296	21	AA124854	Arabidopsis thalia
62	7	1.4	301	21	AA124853	Arabidopsis thalia
63	7	1.4	308	22	AA148546	Propionibacterium
64	7	1.4	312	22	AA181921	S. epidermidis ope
65	7	1.4	312	22	AA135466	Escherichia coli E
66	7	1.4	312	22	AA135471	Escherichia coli E
67	7	1.4	313	21	AA181138	Bordetella pertuss
68	7	1.4	320	23	ABP39594	Staphylococcus epi
69	7	1.4	327	20	AA105529	Soybean vestitone
70	7	1.4	327	21	AA197832	Adenovirus 5 recom
71	7	1.4	329	23	AA187041	Adenovirus 5 recom
72	7	1.4	329	23	AA187042	37KD protein #2.
73	7	1.4	331	21	AA136831	Streptococcus poly
74	7	1.4	341	23	ABP27788	MBP protein. Hom
75	7	1.4	344	23	ABP43708	Arabidopsis thalia
76	7	1.4	365	21	AA109073	Arabidopsis thalia
77	7	1.4	365	21	AA109073	Arabidopsis thalia
78	7	1.4	373	23	ABP54724	Lactococcus lactis
79	7	1.4	375	23	ABP43792	MBP protein #2.
80	7	1.4	376	22	ABP60164	Drosophila melanog
81	7	1.4	377	22	AA193690	Human polypeptide,
82	7	1.4	387	23	ABP53661	Lactococcus lactis

83	7	1.4	436	17	AAR95636	Carilage-derived
84	7	1.4	438	18	AAW18209	Staphylococcus aur
85	7	1.4	439	22	AAU00837	S. aureus Coenzym
86	7	1.4	440	21	AAZ20592	Aspergillus terreu
87	7	1.4	440	21	AAI69545	Aspergillus terreu
88	7	1.4	451	21	AAI12576	NDO related comple
89	7	1.4	466	19	AAW84357	Aspergillus terreu
90	7	1.4	472	22	ABG23847	Novel human diagno
91	7	1.4	478	21	AAI35988	Sorbitol dehydroge
92	7	1.4	486	23	ABR05347	Human nuclear hoim
93	7	1.4	489	22	ABR67879	Drosophila melanog
94	7	1.4	511	22	ABR64688	Drosophila melanog
95	7	1.4	512	10	AAZ94880	FUS 1 promoter inc
96	7	1.4	512	12	AAI49098	ORF from BIK1-Fusi
97	7	1.4	512	12	AAI49098	Fusi protein. Sac
98	7	1.4	516	24	ABU01580	S. pneumoniae type
99	7	1.4	519	24	ABP71695	Human estrogen-rel
100	7	1.4	521	9	AAE09030	Sequence of human
101	7	1.4	525	21	AAZ09072	Arabidopsis thalia
102	7	1.4	525	21	AAZ09072	Arabidopsis thalia
103	7	1.4	525	23	ABR93308	Herbicidally activ
104	7	1.4	534	20	AAH82789	Human IRS-p53h pro
105	7	1.4	540	21	AAH43056	Human ORF ORF2820
106	7	1.4	544	21	AAZ57302	G. lamblia pep pro
107	7	1.4	549	21	AAZ09071	Arabidopsis thalia
108	7	1.4	549	21	AAZ53092	Arabidopsis thalia
109	7	1.4	551	20	AAZ73938	Human prostate tum
110	7	1.4	552	22	AAH39411	Human polypeptide
111	7	1.4	556	22	AAH41187	Human polypeptide
112	7	1.4	557	22	ABR11470	Drosophila melanog
113	7	1.4	565	24	ABU11766	Human MDR polypep
114	7	1.4	567	18	AAW55703	H. pylori ORF 07ee
115	7	1.4	587	22	AAH94824	Human-protein sequ
116	7	1.4	587	23	AAE15554	Human secreted pro
117	7	1.4	611	22	AAU36518	Pseudomonas aerugi
118	7	1.4	617	18	AAW27247	Pseudomonas fluore
119	7	1.4	659	23	AAU98892	Human protease PR
120	7	1.4	683	22	ABG30138	Novel human diagno
121	7	1.4	687	22	ABR64229	Drosophila melanog
122	7	1.4	700	21	AAZ75589	Neisseria gonorrhoe
123	7	1.4	700	21	ABP77608	N. gonorrhoeae am
124	7	1.4	747	23	ABR53579	Lactococcus lactis
125	7	1.4	792	23	AAU83419	Novel Lactobacillu
126	7	1.4	844	19	AAW98721	H. pylori GHP0 815
127	7	1.4	859	17	AAW07872	Aluminium resistan
128	7	1.4	886	18	AAH89787	Staphylococcus aur
129	7	1.4	891	22	AAE00586	Human nuclear cell
130	7	1.4	895	24	ABP56877	Pathogen specific
131	7	1.4	909	24	ABP56877	Staphylococcus epi
132	7	1.4	946	22	AAE00582	Human nuclear cell
133	7	1.4	962	21	AAZ49944	Human potassium io
134	7	1.4	988	22	AAE01043	Human ether a gogo
135	7	1.4	988	22	AAH31714	A human alpha-subu
136	7	1.4	988	23	ABG70999	Human transporter
137	7	1.4	988	23	ABR76165	Human potassium ch
138	7	1.4	988	23	AAO14214	Human transporter
139	7	1.4	989	21	AAZ49945	Human potassium io
140	7	1.4	1018	22	AAE00581	Human cell adhesio
141	7	1.4	1047	22	AAW78507	Human protein SEQ
142	7	1.4	1047	22	AAW79491	Human protein SEQ
143	7	1.4	1095	24	ABJ25381	Aspergillus fumiga
144	7	1.4	1102	23	ABR93939	Herbicidally activ
145	7	1.4	1174	22	ABR65304	Drosophila melanog
146	7	1.4	1187	20	AAZ05841	Banana ripening tr
147	7	1.4	1204	21	AAZ48838	Arabidopsis thalia
148	7	1.4	1216	21	AAZ48837	Arabidopsis thalia
149	7	1.4	1219	22	ABR65388	Drosophila melanog
150	7	1.4	1239	22	ABR61801	Drosophila melanog
151	7	1.4	1248	21	AAZ48836	Arabidopsis thalia
152	7	1.4	1277	21	ABZ75981	Aspergillus fumiga
153	7	1.4	1366	22	ABR66921	Drosophila melanog
154	7	1.4	1447	21	AAH36514	Candida albicans C
155	7	1.4	1447	23	ABP73835	Candida albicans e
156	7	1.4	1461	24	ABP57442	Mycobacterium tube
157	7	1.4	1568	23	ABG97467	Human NAAP1, from
158	7	1.4	1594	22	ABR63688	Drosophila melanog
159	7	1.4	1576	24	ABJ25188	Aspergillus fumiga
160	7	1.4	1891	19	AAW52844	Aspergillus medi
161	7	1.4	1941	23	ABR65509	Blitodacterium lo
162	7	1.4	2705	22	AAH90772	Human shear stress
163	7	1.4	3080	14	AAH35081	ZYVW polyprotein..
164	7	1.4	3413	19	AAW52849	A. mediterranei ri
165	7	1.4	4498	22	ABR58595	Drosophila melanog
166	7	1.4	7068	22	AAE10142	Streptomyces nous
167	7	1.4	9477	22	AAE10144	Streptomyces nous
168	7	1.4	9477	22	AAE10129	Streptomyces nous
169	6	1.2	11096	6	AAZ55777	PDZ motif sequence
170	6	1.2	6	22	AAZ55769	Neurexin C-termina
171	6	1.2	6	22	AAZ58019	Neurexin C-termina
172	6	1.2	6	22	ABU05275	T-cell surface rec
173	6	1.2	6	23	ABP63503	PDZ motif (PL) pep
174	6	1.2	6	23	AAW64500	Neurotoxic beta-am
175	6	1.2	7	19	AAZ55778	PDZ motif sequence
176	6	1.2	7	22	AAZ57610	Neurexin C-termina
177	6	1.2	7	22	AAZ58020	Neurexin C-termina
178	6	1.2	7	23	ABU05276	T-cell surface rec
179	6	1.2	7	23	ABP63504	PDZ motif (PL) pep
180	6	1.2	7	23	AAU99458	Dansyl-labeled be
181	6	1.2	8	20	AAZ06067	Human cancer anti
182	6	1.2	8	22	AAZ55779	PDZ motif sequence
183	6	1.2	8	22	AAZ57611	Neurexin C-termina
184	6	1.2	8	22	AAZ58021	Neurexin C-termina
185	6	1.2	8	23	ABU05277	T-cell surface rec
186	6	1.2	8	23	ABP63505	PDZ motif (PL) pep
187	6	1.2	9	15	AAZ55742	Protein-kinase inh
188	6	1.2	9	20	AAZ06046	Human cancer anti
189	6	1.2	9	20	AAZ05868	Human cancer anti
190	6	1.2	9	21	AAZ79750	Human cancer anti
191	6	1.2	9	24	ABP74471	Human NY-ESO-1 deri
192	6	1.2	10	19	AAH30675	Brissica napus acy
193	6	1.2	10	20	AAZ06069	Human cancer anti
194	6	1.2	10	20	AAZ06070	Human cancer anti
195	6	1.2	10	20	AAZ06071	Human cancer anti
196	6	1.2	10	20	AAZ06072	Human cancer anti
197	6	1.2	10	20	AAZ06015	Human cancer anti
198	6	1.2	10	20	AAZ05969	Human cancer anti
199	6	1.2	10	20	AAZ05970	Human cancer anti
200	6	1.2	10	20	AAZ05971	Human cancer anti
201	6	1.2	10	20	AAZ05972	Human cancer anti
202	6	1.2	10	20	AAZ05973	Human cancer anti
203	6	1.2	10	20	AAZ05974	Human cancer anti
204	6	1.2	10	20	AAZ05976	Human cancer anti
205	6	1.2	10	20	AAZ05977	Human cancer anti
206	6	1.2	10	22	AAZ07734	Human MHC class I
207	6	1.2	10	22	AAZ31331	Exemplary antigen
208	6	1.2	10	23	ABG66802	Tumour antigen NY-
209	6	1.2	11	20	AAZ06066	Human cancer anti
210	6	1.2	11	20	AAZ06068	Human cancer anti
211	6	1.2	11	20	AAZ05967	Human cancer anti
212	6	1.2	11	21	AAZ99291	Human MHC class I
213	6	1.2	11	22	AAZ07733	Human MHC class I
214	6	1.2	11	22	AAZ04942	Nuclear Dbp2-relat
215	6	1.2	11	22	AAZ04943	Proteome analysis
216	6	1.2	11	24	ABP74798	MHC binding peptid
217	6	1.2	12	13	AAZ21768	MHC binding peptid
218	6	1.2	12	14	AAZ34801	Major histocompati
219	6	1.2	12	20	AAZ06065	Human cancer anti
220	6	1.2	12	21	AAZ99269	HLA class II bindi
221	6	1.2	12	23	AAU87915	PDZ domain binding
222	6	1.2	12	23	AAU87916	PDZ domain binding
223	6	1.2	13	13	AAZ21769	MHC binding peptid
224	6	1.2	13	14	AAZ34807	Major histocompati
225	6	1.2	13	14	AAZ34808	Major histocompati
226	6	1.2	13	14	AAZ34809	Major histocompati
227	6	1.2	13	14	AAZ34810	Major histocompati
228	6	1.2	13	14	AAZ34811	Major histocompati



229	6	1.2	13	14	AA34812	Major histocompati	302	6	1.2	34	22	ABG59387	Human liver peptid
230	6	1.2	13	14	AA34813	Major histocompati	303	6	1.2	34	22	ABG23702	Novel human diago
231	6	1.2	13	14	AA34814	Major histocompati	304	6	1.2	34	22	ABB4008	Peptide #7594 enco
232	6	1.2	13	14	AA34815	Major histocompati	305	6	1.2	34	22	ABB44000	Peptide #11506 enc
233	6	1.2	13	14	AA34816	Major histocompati	306	6	1.2	34	22	ABB24572	Protein #6571 enco
234	6	1.2	13	14	AA34817	Major histocompati	307	6	1.2	34	22	ABB26908	Protein #8907 enco
235	6	1.2	13	14	AA34818	Major histocompati	308	6	1.2	34	22	AA660845	Human brain expres
236	6	1.2	13	14	AA34819	Major histocompati	309	6	1.2	34	22	AA665014	Human brain expres
237	6	1.2	13	14	AA34834	Major histocompati	310	6	1.2	34	22	AA673528	Human bone marrow
238	6	1.2	13	14	AA34836	Major histocompati	311	6	1.2	34	22	AA677731	Human bone marrow
239	6	1.2	13	14	AA34837	Major histocompati	312	6	1.2	34	22	AA620001	Peptide #6435 enco
240	6	1.2	13	14	AA34838	Major histocompati	313	6	1.2	34	22	AA621640	Peptide #8074 enco
241	6	1.2	13	14	AA34839	Major histocompati	314	6	1.2	34	22	AA633722	Peptide #7759 enco
242	6	1.2	13	14	AA34820	Major histocompati	315	6	1.2	34	22	AA637945	Peptide #11982 enc
243	6	1.2	13	14	AA34821	Major histocompati	316	6	1.2	34	22	AA699761	ERA binding domain
244	6	1.2	13	14	AA34822	Major histocompati	317	6	1.2	34	23	ABG43389	Human peptide enco
245	6	1.2	13	14	AA34823	Major histocompati	318	6	1.2	34	23	ABG46761	Human peptide enco
246	6	1.2	13	14	AA34824	Major histocompati	319	6	1.2	37	22	ABB02689	Vital core polypep
247	6	1.2	13	14	AA34825	Major histocompati	320	6	1.2	38	22	AA699869	ERA binding domain
248	6	1.2	13	14	AA34826	Major histocompati	321	6	1.2	38	23	ABB82354	Pufferfish urococt
249	6	1.2	13	14	AA34827	Major histocompati	322	6	1.2	38	23	ABP25992	Streptococcus poly
250	6	1.2	13	14	AA34828	Major histocompati	323	6	1.2	39	21	AA698830	Core polypeptide f
251	6	1.2	13	14	AA34829	Major histocompati	324	6	1.2	39	22	ABB01238	Vital Dp18/107-1i
252	6	1.2	13	14	AA34830	Major histocompati	325	6	1.2	39	22	AA613784	Dp178-11ke/Dp107-1
253	6	1.2	13	14	AA34831	Major histocompati	326	6	1.2	39	22	AA678231	Core polypeptide T
254	6	1.2	13	14	AA34832	Major histocompati	327	6	1.2	39	23	ABP53558	Plasmod PRSET Sfil
255	6	1.2	13	14	AA34833	Major histocompati	328	6	1.2	39	23	ABP53560	Plasmod PRSET Sfil
256	6	1.2	13	18	AAW22116	Padre (pan-DR bind	329	6	1.2	39	23	ABP53562	Plasmod PRSET Sfil
257	6	1.2	13	19	AAW50121	Pan DR binding pep	330	6	1.2	40	22	AA658234	Human brain expres
258	6	1.2	13	20	AA606064	Human cancer antiq	331	6	1.2	40	22	AA699833	ERA binding domain
259	6	1.2	13	21	AA626807	Phosphoryl group a	332	6	1.2	42	22	AA699788	ERA binding domain
260	6	1.2	13	21	AA699270	HLA class II bindi	333	6	1.2	42	22	AA699797	ERA binding domain
261	6	1.2	13	23	ABP54999	Cyclin dependent p	334	6	1.2	43	20	AA682294	Dynamn 41.2 ZELan
262	6	1.2	13	24	ABP59793	Peptide used as ph	335	6	1.2	43	23	AA615500	Porphyrin yezoensis
263	6	1.2	13	24	ABP56628	Cyclin dependent p	336	6	1.2	44	22	AA608629	Human polypeptide
264	6	1.2	13	24	ABP59484	Phosphoryl group a	337	6	1.2	46	22	ABG01742	Novel human diago
265	6	1.2	13	24	ABP55027	Cyclin dependent p	338	6	1.2	46	22	AA699806	ERA binding domain
266	6	1.2	14	14	AA34835	Major histocompati	339	6	1.2	46	22	AA699854	ERA binding domain
267	6	1.2	14	18	AAW15095	Anti-HIV p24 antib	340	6	1.2	47	21	AA609429	Hepatitis GB virus
268	6	1.2	14	20	AA605986	Human cancer antiq	341	6	1.2	47	22	AA699791	ERA binding domain
269	6	1.2	14	22	ABP25107	Human MHC peptide	342	6	1.2	47	22	AA699800	ERA binding domain
270	6	1.2	14	22	AAW00229	Human eph fragment	343	6	1.2	48	20	AA660201	Human endometrium
271	6	1.2	14	22	AAW00231	Human eph fragment	344	6	1.2	49	22	ABG50143	Human liver peptid
272	6	1.2	14	22	AA603837	Hepatitis C virus	345	6	1.2	49	22	ABB30100	Peptide #2751 enco
273	6	1.2	14	23	ABJ09705	Hepatitis B virus	346	6	1.2	49	22	ABB35271	Peptide #2777 enco
274	6	1.2	15	20	AA605978	Human cancer antiq	347	6	1.2	49	22	ABB20713	Protein #2712 enco
275	6	1.2	19	22	AA655819	PL peptide #30. S	348	6	1.2	49	22	AA656102	Human brain expres
276	6	1.2	19	22	AA657681	Neurexin PL peptid	349	6	1.2	49	22	AA668472	Human bone marrow
277	6	1.2	19	22	AA658091	Neurexin PL peptid	350	6	1.2	49	22	AA616281	Peptide #2715 enco
278	6	1.2	19	23	ABJ05211	C-terminal peptide	351	6	1.2	49	22	AA628768	Peptide #2805 enco
279	6	1.2	19	23	ABP63315	Neurexin PL peptid	352	6	1.2	49	22	AA699770	ERA binding domain
280	6	1.2	20	20	AA605979	Human cancer antiq	353	6	1.2	49	22	AA699899	ERA binding domain
281	6	1.2	20	20	AA682295	Dynamn 41.20 ZELa	354	6	1.2	49	22	AA604014	Peptide #2696 enco
282	6	1.2	22	24	AA616223	Agrobacterium tune	355	6	1.2	49	23	ABG38053	Human peptide enco
283	6	1.2	23	20	AA682310	Transport peptide	356	6	1.2	50	22	AAU42299	Propionibacterium
284	6	1.2	23	22	AA698276	Anti-A33 antigen i	357	6	1.2	50	22	AA699842	ERA binding domain
285	6	1.2	23	22	AA698277	Anti-A33 antigen i	358	6	1.2	51	22	AA606820	Human foetal prote
286	6	1.2	23	22	AA698278	Anti-A33 antigen i	359	6	1.2	51	22	AA606825	Human foetal prote
287	6	1.2	23	22	AA697656	A33 antigen bindin	360	6	1.2	52	21	AA64748	Human 5' EST relat
288	6	1.2	23	22	AA697657	A33 antigen bindin	361	6	1.2	52	22	AA699764	ERA binding domain
289	6	1.2	23	22	AA697658	A33 antigen bindin	362	6	1.2	53	19	AAW38647	S. pneumoniae ATP
290	6	1.2	25	22	ABG54356	Human liver peptid	363	6	1.2	53	22	AAU42426	Propionibacterium
291	6	1.2	25	22	ABP39377	Peptide #6883 enco	364	6	1.2	53	22	AAU56774	Propionibacterium
292	6	1.2	25	22	ABB24164	Protein #6163 enco	365	6	1.2	54	22	AA699848	ERA binding domain
293	6	1.2	25	22	AA660055	Human brain expres	366	6	1.2	54	22	AA677023	Mutant GHR exon 7
294	6	1.2	25	22	AAW72659	Human bone marrow	367	6	1.2	55	16	AA677020	Mutant GHR exon 7
295	6	1.2	25	22	AA619670	Peptide #6104 enco	368	6	1.2	55	16	AA677021	Mutant GHR exon 7
296	6	1.2	25	22	AA632892	Peptide #6929 enco	369	6	1.2	55	16	AA677022	Human GHR exon 7 p
297	6	1.2	25	23	ABG62600	Bubacterial DnaE1	370	6	1.2	55	18	AAW44242	Growth hormone rec
298	6	1.2	25	23	ABG42483	Human peptide enco	371	6	1.2	55	18	AAW44243	Growth hormone rec
299	6	1.2	30	23	AA685105	Human NIN501a segm	372	6	1.2	55	18	AAW44244	Growth hormone rec
300	6	1.2	31	20	AAW70858	Human Zcyto5 epit	373	6	1.2	55	18	AAW44245	Growth hormone rec
301	6	1.2	34	22	ABG55255	Human liver peptid	374	6	1.2	55	19	AAW70558	Growth hormone rec

375	6	1.2	55	19	AAW70559	Growth hormone rec	448	6	1.2	70	22	AAW19372	Peptide #5806 enco
376	6	1.2	55	19	AAW70560	Growth hormone rec	449	6	1.2	70	22	AAW32224	Peptide #6261 enco
377	6	1.2	55	19	AAW70561	Growth hormone rec	450	6	1.2	70	22	ABG41760	Human peptide enco
378	6	1.2	55	22	ABG08147	Novel human diagno	451	6	1.2	71	22	AAU44830	Proionibacterium
379	6	1.2	55	22	ABB17780	Human nervous syst	452	6	1.2	71	22	AAG99633	ERA binding domain
380	6	1.2	55	23	ABP05003	Human ORFX protein	453	6	1.2	71	22	AAAG9634	ERA binding domain
381	6	1.2	56	22	ABG29048	Novel human diagno	454	6	1.2	71	23	AAU99186	Target molecule hu
382	6	1.2	57	22	ABG91744	C glutamicum prote	455	6	1.2	72	22	ABG59207	Human liver peptid
383	6	1.2	57	23	ABP32584	Human ORF1557 prot	456	6	1.2	72	22	AAU49195	Proionibacterium
384	6	1.2	58	22	ABB96408	Human testicular a	457	6	1.2	72	22	ABG08844	Novel human diagno
385	6	1.2	58	22	AAU47762	Proionibacterium	458	6	1.2	72	22	AAW64808	Human brain expres
386	6	1.2	58	22	AAW5877	Human reproductiv	459	6	1.2	72	22	AAW77564	Human bone marrow
387	6	1.2	58	22	AAW92725	Human digestive sy	460	6	1.2	72	22	AAO02535	Human polyptide
388	6	1.2	58	22	AAAG9860	ERA binding domain	461	6	1.2	72	22	AAW37746	Peptide #11783 enc
389	6	1.2	58	22	AAAG9866	ERA binding domain	462	6	1.2	72	23	ABG46592	Human peptide enco
390	6	1.2	59	21	AAAG3846	Arabidopsis thalia	463	6	1.2	73	21	AAW52618	Helicobacter pylor
391	6	1.2	59	22	ABG52518	Human liver peptid	464	6	1.2	73	22	ABW96455	Novel testicular a
392	6	1.2	59	22	ABG58236	Human liver peptid	465	6	1.2	73	22	ABG25287	Novel human diagno
393	6	1.2	59	22	AAU56677	Proionibacterium	466	6	1.2	73	22	AAW95924	Human reproductiv
394	6	1.2	59	22	ABB42830	Peptide #10336 enc	467	6	1.2	73	23	ABU51054	Helicobacter pylor
395	6	1.2	59	22	ABB26111	Protein #8110 enco	468	6	1.2	74	13	AAW30165	UGT1 Exon 4 produc
396	6	1.2	59	22	AAW63732	Human brain expres	469	6	1.2	74	21	AAW12471	Zea mays protein f
397	6	1.2	59	22	AAW76536	Human bone marrow	470	6	1.2	74	22	AAW54587	Proionibacterium
398	6	1.2	59	22	AAW20926	Peptide #7360 enco	471	6	1.2	75	16	AAW71130	SH2 domain of cyto
399	6	1.2	59	22	AAW36643	Peptide #10680 enc	472	6	1.2	75	21	AAW34737	Human secreted pro
400	6	1.2	59	22	AAAG9662	ERA binding domain	473	6	1.2	75	21	AAW01935	Human secreted pro
401	6	1.2	59	22	AAAG99704	ERA binding domain	474	6	1.2	75	21	AAW87587	Aeromonas caviae a
402	6	1.2	59	22	AAAG99719	ERA binding domain	475	6	1.2	75	22	ABG09773	Novel human diagno
403	6	1.2	59	22	AAAG99725	ERA binding domain	476	6	1.2	75	23	ABP04130	Human ORFX protein
404	6	1.2	59	22	AAAG99812	ERA binding domain	477	6	1.2	76	23	ABP08002	Human ORFX protein
405	6	1.2	59	22	AAAG99818	ERA binding domain	478	6	1.2	77	16	AAW82858	Human ORFX protein
406	6	1.2	59	22	AAAG99890	ERA binding domain	479	6	1.2	77	23	ABW81200	Fragment of alpha-
407	6	1.2	59	23	ABG45789	Human peptide enco	480	6	1.2	78	22	AAU55449	E. coli Z-ring int
408	6	1.2	60	22	AAAG99661	ERA binding domain	481	6	1.2	78	17	AAW90781	Proionibacterium
409	6	1.2	60	22	AAAG99663	ERA binding domain	482	6	1.2	79	23	ABP31043	Reverse-FRAME anti
410	6	1.2	60	22	AAAG99713	ERA binding domain	483	6	1.2	79	23	ABP05175	Human ORF16 protei
411	6	1.2	60	22	AAAG99731	ERA binding domain	484	6	1.2	80	17	AAW90788	Reverse-FRAME anti
412	6	1.2	60	22	AAAG99773	ERA binding domain	485	6	1.2	80	17	AAW90787	Reverse-FRAME anti
413	6	1.2	60	22	AAAG99827	ERA binding domain	486	6	1.2	80	17	AAW90786	Reverse-FRAME anti
414	6	1.2	60	22	AAAG99845	ERA binding domain	487	6	1.2	80	17	AAW90785	Reverse-FRAME anti
415	6	1.2	60	23	ABP42201	Human ovarian anti	488	6	1.2	80	17	AAW90784	Reverse-FRAME anti
416	6	1.2	60	24	ABP97608	Amino acid sequenc	489	6	1.2	80	17	AAW90783	Reverse-FRAME anti
417	6	1.2	62	22	ABG19857	Novel human diagno	490	6	1.2	80	17	AAW90782	Reverse-FRAME anti
418	6	1.2	63	22	ABG20392	Novel human diagno	491	6	1.2	80	22	ABG50283	Human liver peptid
419	6	1.2	63	22	ABG26590	Novel haematologic	492	6	1.2	80	22	AAW66913	Proionibacterium
420	6	1.2	64	22	AAW81428	Human haematologic	493	6	1.2	80	22	ABW30255	Peptide #2906 enco
421	6	1.2	64	22	AAW81692	Human haematologic	494	6	1.2	80	22	ABW35421	Peptide #2927 enco
422	6	1.2	64	23	ABP34873	Human kinase-like	495	6	1.2	80	22	ABW20858	Protein #2857 enco
423	6	1.2	65	20	AAV23796	A gyrase protein s	496	6	1.2	80	22	ABW78220	Human mitochondria
424	6	1.2	65	22	AAU41751	Proionibacterium	497	6	1.2	80	22	AAW56245	Human brain expres
425	6	1.2	65	22	AAU51508	Proionibacterium	498	6	1.2	80	22	AAW68623	Human bone marrow
426	6	1.2	65	22	AAW65653	Human brain expres	499	6	1.2	80	22	AAW16430	Peptide #2864 enco
427	6	1.2	65	22	AAW78269	Human bone marrow	500	6	1.2	80	22	AAW28929	Peptide #2966 enco
428	6	1.2	65	22	AAO05339	Human polyptide	501	6	1.2	80	22	AAW04164	Peptide #2846 enco
429	6	1.2	65	23	ABG47295	Human peptide enco	502	6	1.2	80	23	ABG38201	Human peptide enco
430	6	1.2	65	23	ABP07279	Human ORFX protein	503	6	1.2	81	17	AAW90789	Reverse-FRAME anti
431	6	1.2	66	21	ABW38011	Human secreted pro	504	6	1.2	81	24	AAW79708	Portion of NM 0033
432	6	1.2	66	22	ABG20253	Novel human diagno	505	6	1.2	81	24	AAW79709	Portion of UB2467
433	6	1.2	66	22	ABW4888	Human secreted pro	506	6	1.2	81	24	AAW79711	Portion of mouse 5
434	6	1.2	66	23	AAO15501	Porphyra yezoensis	507	6	1.2	81	24	AAW79712	Portion of US2433
435	6	1.2	68	21	AAAG12472	Zea mays protein f	508	6	1.2	81	24	AAW79713	Portion of US2824
436	6	1.2	68	22	AAU31126	Proionibacterium	509	6	1.2	82	22	AAU56643	Proionibacterium
437	6	1.2	69	21	AAW40541	Human ORFX ORF305	510	6	1.2	82	22	AAW64588	Proionibacterium
438	6	1.2	69	21	AAW37128	Arabidopsis thalia	511	6	1.2	82	22	AAW66340	Proionibacterium
439	6	1.2	69	23	ABW3698	Human gastric canc	512	6	1.2	83	22	ABW41733	Peptide #2939 enco
440	6	1.2	69	23	ABP04101	Human ORFX protein	513	6	1.2	83	22	AAW62606	Human brain expres
441	6	1.2	69	23	ABP05881	Human ORFX protein	514	6	1.2	83	22	AAO03013	Human polyptide
442	6	1.2	70	22	ABG53629	Human liver peptid	515	6	1.2	83	22	AAW35528	Peptide #9565 enco
443	6	1.2	70	22	ABW38751	Peptide #6257 enco	516	6	1.2	83	23	ABP01772	Human ORFX protein
444	6	1.2	70	22	ABW37397	Protein #5796 enco	517	6	1.2	84	22	ABW02717	Novel human diagno
445	6	1.2	70	22	AAW59390	Human brain expres	518	6	1.2	85	22	AAW69450	Human purified sec
446	6	1.2	70	22	AAW71945	Human bone marrow	519	6	1.2	85	22	AAW65806	RP55-like protein
447	6	1.2	70	22	AAO08876	Human polyptide	520	6	1.2	85	23	AAO15498	Porphyra yezoensis

521	6	1.2	86	22	AAM91998	Human digestive sy	594	6	1.2	109	21	AAG34350	Zea mays protein f
522	6	1.2	86	23	ABP64976	Human protein SEQ	595	6	1.2	109	22	AAO09868	Human polypeptide
523	6	1.2	87	22	AG99699	ERA binding domain	596	6	1.2	109	22	AAB98230	Rabbit anti-A33 an
524	6	1.2	88	22	ABG23994	Novel human diago	597	6	1.2	109	22	AAB98231	Rabbit anti-A33 an
525	6	1.2	88	22	AAO08786	Human polypeptide	598	6	1.2	109	22	AAB98232	Rabbit anti-A33 an
526	6	1.2	89	24	AAO2648	Maize basel layer	599	6	1.2	109	22	AAB75149	Rabbit anti-A33 an
527	6	1.2	89	24	AAO2649	Maize basel layer	600	6	1.2	109	22	AAB75150	Rabbit anti A33 an
528	6	1.2	90	22	AAO2646	Maize BTU-2 prote	601	6	1.2	110	15	AAR56390	Extracellular hGR
529	6	1.2	90	24	AAO2647	Maize protein, SEQ	602	6	1.2	110	23	AAO15499	Porphyra yezoensis
530	6	1.2	90	24	AAO2647	Maize protein, SEQ	603	6	1.2	110	23	AAO15499	Porphyra yezoensis
531	6	1.2	91	21	AAO18124	Arabidopsis thalia	604	6	1.2	111	21	AAG54944	Arabidopsis thalia
532	6	1.2	91	21	AAO18124	Arabidopsis thalia	605	6	1.2	111	21	AAG54944	Arabidopsis thalia
533	6	1.2	91	21	AAO18124	Arabidopsis thalia	606	6	1.2	112	15	AAB56391	Extracellular hGR
534	6	1.2	91	21	AAO18124	Arabidopsis thalia	607	6	1.2	112	15	AAB56391	Extracellular hGR
535	6	1.2	92	23	ABP00953	Human ORFX protein	608	6	1.2	113	23	ABP00953	Human ORFX protein
536	6	1.2	92	23	ABP00953	Human ORFX protein	609	6	1.2	113	23	ABP00953	Human ORFX protein
537	6	1.2	94	22	ABP19735	Novel human diago	610	6	1.2	115	21	AAO00563	Zea mays protein f
538	6	1.2	94	22	ABP19735	Novel human diago	611	6	1.2	115	22	AAO02810	Human polypeptide
539	6	1.2	95	21	AAO18124	Arabidopsis thalia	612	6	1.2	115	22	AAO02810	Human polypeptide
540	6	1.2	95	21	AAO18124	Arabidopsis thalia	613	6	1.2	116	21	AAG33958	Arabidopsis thalia
541	6	1.2	95	21	AAO18124	Arabidopsis thalia	614	6	1.2	116	21	AAG33958	Arabidopsis thalia
542	6	1.2	95	21	AAO18124	Arabidopsis thalia	615	6	1.2	116	22	ABG02936	Novel human diago
543	6	1.2	95	22	ABP16615	Human nervous syst	616	6	1.2	117	16	AAM82569	Wheat glycoengin c
544	6	1.2	95	22	ABP16615	Human nervous syst	617	6	1.2	117	22	AAU35520	Haemophilus influe
545	6	1.2	96	22	ABP47811	SMRV protein relat	618	6	1.2	117	22	AAU35520	Haemophilus influe
546	6	1.2	96	22	ABP47811	SMRV protein relat	619	6	1.2	117	22	AAU35520	Haemophilus influe
547	6	1.2	97	22	ABG21324	Novel human diago	620	6	1.2	118	22	ABG20410	Novel human diago
548	6	1.2	97	23	ABP47808	MPV protein relat	621	6	1.2	118	22	ABG20410	Novel human diago
549	6	1.2	97	23	ABP47808	MPV protein relat	622	6	1.2	119	22	AAO06904	Human polypeptide
550	6	1.2	97	23	ABP47808	MPV protein relat	623	6	1.2	120	22	ABP69277	Human polypeptide
551	6	1.2	97	23	ABP47808	MPV protein relat	624	6	1.2	121	21	AAO10885	Drosophila melanog
552	6	1.2	97	23	ABP47808	MPV protein relat	625	6	1.2	121	21	AAO10885	Drosophila melanog
553	6	1.2	98	21	AAO10885	Human secreted pro	626	6	1.2	121	21	AAO10885	Human secreted pro
554	6	1.2	98	21	AAO10885	Human secreted pro	627	6	1.2	121	21	AAO10885	Human secreted pro
555	6	1.2	98	22	AAU48448	Propionibacterium	628	6	1.2	121	22	ABG06733	Novel human diago
556	6	1.2	99	21	AAO1554	Arabidopsis thalia	629	6	1.2	121	22	ABG29522	Novel human diago
557	6	1.2	99	22	ABG22584	Novel human diago	630	6	1.2	122	23	ABP01617	Human ORFX protein
558	6	1.2	99	22	ABG22584	Novel human diago	631	6	1.2	123	21	AAV93964	Mouse acid sequenc
559	6	1.2	99	22	AAU33019	Novel human secret	632	6	1.2	123	23	AAE26441	Mouso TANGO 219 pr
560	6	1.2	100	20	AAU29205	Amino acid sequenc	633	6	1.2	124	21	AAG24036	Arabidopsis thalia
561	6	1.2	100	22	AAU60043	Propionibacterium	634	6	1.2	124	23	ABU10908	Human secreted pro
562	6	1.2	100	22	AAU60043	Propionibacterium	635	6	1.2	126	14	AAU50054	Human secreted pro
563	6	1.2	101	21	AAO59816	Arabidopsis thalia	636	6	1.2	127	21	AAO59816	Arabidopsis thalia
564	6	1.2	101	22	AAU59816	Arabidopsis thalia	637	6	1.2	127	22	AAU59816	Arabidopsis thalia
565	6	1.2	101	22	AAU59816	Arabidopsis thalia	638	6	1.2	127	22	AAU59816	Arabidopsis thalia
566	6	1.2	101	23	ABP64398	Corynebacterium gl	639	6	1.2	127	23	ABG96450	Human polypeptide
567	6	1.2	101	23	ABP64398	Corynebacterium gl	640	6	1.2	128	21	ABP39172	Human ovarian canc
568	6	1.2	101	23	ABP64398	Corynebacterium gl	641	6	1.2	128	21	ABP39172	Human ovarian canc
569	6	1.2	101	24	ABP05739	Human ORFX protein	642	6	1.2	128	23	ABP69240	Human polypeptide
570	6	1.2	102	21	AAO10338	Human ORFX ORF329	643	6	1.2	129	21	AAO10338	Human polypeptide
571	6	1.2	102	21	AAO10338	Human ORFX ORF329	644	6	1.2	129	21	AAO10338	Human polypeptide
572	6	1.2	102	21	AAO10338	Human ORFX ORF329	645	6	1.2	129	22	AAU33662	Pseudomonas aerugi
573	6	1.2	102	21	AAO10338	Human ORFX ORF329	646	6	1.2	129	22	AAU33662	Pseudomonas aerugi
574	6	1.2	102	22	AAO10338	Human ORFX ORF329	647	6	1.2	129	22	AAU33662	Pseudomonas aerugi
575	6	1.2	102	23	ABP32681	Human polypeptide	648	6	1.2	129	24	ABP81256	Human protein sequ
576	6	1.2	103	24	ABP58239	Human ORF1664 prot	649	6	1.2	130	22	AAU33503	Arabidopsis thalia
577	6	1.2	105	23	ABP64328	Human ovary-specif	650	6	1.2	130	22	AAU33503	Arabidopsis thalia
578	6	1.2	106	17	ABP90791	Human ORF698, Hom	651	6	1.2	130	22	AAU33503	Arabidopsis thalia
579	6	1.2	106	21	AAO40735	Reverse-Frame anti	652	6	1.2	130	22	AAU33503	Arabidopsis thalia
580	6	1.2	106	21	AAO40735	Reverse-Frame anti	653	6	1.2	130	23	ABP25699	Streptococcus poly
581	6	1.2	107	21	ABP02646	Human ORFX protein	654	6	1.2	130	24	ABP02646	Streptococcus poly
582	6	1.2	107	22	AAO18390	Human immune/haema	655	6	1.2	131	21	AAO18390	Human immune/haema
583	6	1.2	107	23	AAO18390	Human immune/haema	656	6	1.2	131	21	AAO18390	Human immune/haema
584	6	1.2	107	23	AAO18390	Human immune/haema	657	6	1.2	131	22	AAU35463	Haemophilus influe
585	6	1.2	107	23	AAO18390	Human immune/haema	658	6	1.2	131	22	AAU35463	Haemophilus influe
586	6	1.2	107	23	AAO18390	Human immune/haema	659	6	1.2	131	23	ABG93340	Human protein sequ
587	6	1.2	107	24	ABP76286	Human GENSERT prote	660	6	1.2	132	22	AAU87477	Novel central nerv
588	6	1.2	108	22	AAU34776	E. coli cellular p	661	6	1.2	132	22	AAU34776	Novel central nerv
589	6	1.2	108	22	AAU34776	E. coli cellular p	662	6	1.2	132	22	AAU34776	Novel central nerv
590	6	1.2	108	22	ABP17577	Human nervous syst	663	6	1.2	132	23	ABP48070	Human novel extrac
591	6	1.2	108	23	ABP29635	Streptococcus poly	664	6	1.2	133	22	ABP48070	Human novel extrac
592	6	1.2	109	19	AAW38664	S. pneumoniae 305	665	6	1.2	133	23	ABP48070	Human novel extrac
593	6	1.2	109	21	AAW38664	S. pneumoniae 305	666	6	1.2	133	24	ABU67323	Novel human diago

Zea mays protein f  
Human polypeptide  
Rabbit anti-A33 an  
Rabbit anti-A33 an  
Rabbit anti-A33 an  
Rabbit anti-A33 an  
Rabbit anti A33 an  
Rabbit anti A33 an  
Extracellular hGR  
Porphyra yezoensis  
Arabidopsis thalia  
Extracellular hGR  
Drosophila melanog  
Human polypeptide  
Human polypeptide  
Human ORFX protein  
Zea mays protein f  
Human polypeptide  
Human polypeptide  
Arabidopsis thalia  
Arabidopsis thalia  
Novel human diago  
Wheat glycoengin c  
Haemophilus influe  
Wheat water stress  
Wheat water stress  
Novel human diago  
Human polypeptide  
Human polypeptide  
Drosophila melanog  
Human secreted pro  
Human secreted pro  
Human secreted pro  
Human secreted pro  
Human secreted pro  
Novel human diago  
Novel human diago  
Human ORFX protein  
Mouse acid sequenc  
Mouso TANGO 219 pr  
Arabidopsis thalia  
Human secreted pro  
IC94.5 fragment.  
Arabidopsis thalia  
Propionibacterium  
Human polypeptide  
Human polypeptide,  
Human ovarian canc  
Human secreted pro  
ERA binding domain  
Human polypeptide  
Arabidopsis thalia  
Arabidopsis thalia  
Pseudomonas aerugi  
Propionibacterium  
Human protein sequ  
Arabidopsis thalia  
Enterococcus faeca  
Streptococcus pneu  
Streptococcus pneu  
Streptococcus pneu  
Streptococcus poly  
Streptococcus pneu  
S. pneumoniae type  
Gene 1 human secre  
Human secreted pro  
Haemophilus influe  
Human protein sequ  
C. albicans BAX-as  
Novel central nerv  
Propionibacterium  
Human novel extrac  
Human polypeptide  
Novel human diago  
Novel GGP synthase  
Corn geranylgerany

667	6	1.2	134	9	AAp81139	Sequence of plant	740	6	1.2	156	22	ABG15620	Novel human diagno
668	6	1.2	134	13	AAr28465	Sequence encoded b	741	6	1.2	157	19	AAW56252	Interleukin-13 bin
669	6	1.2	134	15	AAr54979	Brassic seed acyl	742	6	1.2	157	22	AAU86952	Human DNA repair a
670	6	1.2	134	16	AAr76695	B. campestris ACP	743	6	1.2	157	22	ABG01094	Novel human diagno
671	6	1.2	134	18	AAW15416	ACP protein encode	744	6	1.2	157	22	ABG53058	Novel human diagno
672	6	1.2	134	19	AAW30527	Brassic campestr	745	6	1.2	157	22	AAU21724	Escherichia coli p
673	6	1.2	134	19	AAW30671	Brassic campestr	746	6	1.2	157	22	AAU21834	Novel human neopla
674	6	1.2	134	21	AAV49927	Partial genomic AC	747	6	1.2	158	22	AAU87210	Novel human nerv
675	6	1.2	134	22	AAAB85932	Amino acid sequenc	748	6	1.2	158	22	AAU87520	Novel central nerv
676	6	1.2	134	23	ABH89266	Human polypeptide	749	6	1.2	158	22	AAU17724	Novel human respir
677	6	1.2	135	21	AAAG57675	Arabidopsis thalia	750	6	1.2	158	22	AAU19727	Human novel extiec
678	6	1.2	136	21	AAAG5074	Arabidopsis thalia	751	6	1.2	158	22	ABP47947	Human polypeptide
679	6	1.2	136	22	AAU58616	Propionibacterium	752	6	1.2	159	21	AAW41784	Human ORFX ORF1548
680	6	1.2	136	22	ABG28035	Novel diagno	753	6	1.2	159	21	AAV71041	Streptococcus pyog
681	6	1.2	136	22	AAAG4105	Human immune/haema	754	6	1.2	161	22	AAW65804	RPP5-like protein
682	6	1.2	136	22	AAAG42362	Human polypeptide	755	6	1.2	161	23	ABP32415	Human transcrip
683	6	1.2	137	21	AAAG16637	Arabidopsis thalia	756	6	1.2	162	23	ABU05775	M. tuberculosis an
684	6	1.2	137	21	AAAG43076	Arabidopsis thalia	757	6	1.2	162	23	ABW53404	Lactococcus lactis
685	6	1.2	137	21	ABH11435	Human C1RP-1 homol	758	6	1.2	163	20	AAV41233	A. chroococcum YH2
686	6	1.2	137	24	ABP79973	N. gonorrhoeae ami	759	6	1.2	163	21	AAW40557	Human ORFX ORF321
687	6	1.2	138	18	AAW55485	H. pylori ORF hpjd	760	6	1.2	163	21	AAW43062	Human ORFX ORF2826
688	6	1.2	138	22	AAW55485	C glutamicum proce	761	6	1.2	163	22	AAW89040	Human immune/haema
689	6	1.2	138	22	AAAG90088	Novel human membra	762	6	1.2	164	21	AAV87867	M. tuberculosis an
690	6	1.2	138	22	AAU01212	Human protein sequ	763	6	1.2	164	22	ABG14120	Novel human diagno
691	6	1.2	138	22	AAAG95866	Streptococcus poly	764	6	1.2	164	22	AAW71392	Human gene 23-enco
692	6	1.2	138	23	ABP26151	Human cancer assoc	765	6	1.2	164	23	ABU51280	Helicobacter pylor
693	6	1.2	140	21	AAW44225	Arabidopsis thalia	766	6	1.2	164	24	ABP80985	N. gonorrhoeae ami
694	6	1.2	140	21	AAAG57674	Human secreted pro	767	6	1.2	165	21	AAW52613	Helicobacter pylor
695	6	1.2	140	21	AAAG01215	Human immune/haema	768	6	1.2	165	21	AAW09965	Arabidopsis thalia
696	6	1.2	140	22	AAW82572	Streptococcus pyog	769	6	1.2	165	21	AAW50453	Arabidopsis thalia
697	6	1.2	141	21	AAW71040	Human transmembran	770	6	1.2	165	23	ABW92888	Herbicideally activ
698	6	1.2	141	22	AAU52928	Propionibacterium	771	6	1.2	166	19	AAW59153	L. lactis rhnb C-t
699	6	1.2	142	22	AAU56689	Corn GGPP synthase	772	6	1.2	166	19	AAW59162	A. lactis MG316 r
700	6	1.2	142	23	AAE25927	Human polypeptide	773	6	1.2	166	21	AAAG10731	Arabidopsis thalia
701	6	1.2	142	23	ABW90131	Corn geranylgerany	774	6	1.2	166	22	AAW63770	Human prostate can
702	6	1.2	142	24	ABW67324	Zea mays protein f	775	6	1.2	166	22	AAW63812	Human prostate can
703	6	1.2	143	21	AAW41131	Novel human diagno	776	6	1.2	166	23	ABP29017	Streptococcus poly
704	6	1.2	143	22	ABG19723	Human polypeptide	777	6	1.2	167	21	AAAG09244	Arabidopsis thalia
705	6	1.2	143	22	AAO02945	Human ORFX protein	778	6	1.2	167	21	AAW15940	E. coli proliferat
706	6	1.2	143	23	ABP01185	Arabidopsis thalia	779	6	1.2	167	21	AAV71046	Streptococcus pyog
707	6	1.2	144	21	AAAG09966	Arabidopsis thalia	780	6	1.2	167	21	AAV41298	Neisseria chimeric
708	6	1.2	144	21	AAAG50454	Arabidopsis thalia	781	6	1.2	167	22	AAW85153	Novel cytochrome p
709	6	1.2	144	22	ABG18607	Novel human diagno	782	6	1.2	167	23	ABP52144	Zebrafish P450RA1-
710	6	1.2	144	22	AAW71040	Human olfactory re	783	6	1.2	167	23	ABW94077	Human secreted pro
711	6	1.2	146	21	AAW24791	Plant SDF encoded	784	6	1.2	168	22	ABG20391	Novel human diagno
712	6	1.2	146	21	AAW24954	Human novel secret	785	6	1.2	170	21	AAW07548	Arabidopsis thalia
713	6	1.2	146	22	AAU16613	Human novel secret	786	6	1.2	170	21	AAW09576	Arabidopsis thalia
714	6	1.2	146	24	ABW55682	Human novel polype	787	6	1.2	170	21	AAW52219	Arabidopsis thalia
715	6	1.2	148	19	AAW98432	H. pylori GHP 447	788	6	1.2	170	21	AAW57728	Arabidopsis thalia
716	6	1.2	148	22	AAO06229	Human polypeptide	789	6	1.2	170	22	ABG20255	Novel human diagno
717	6	1.2	149	22	AAW64547	Human ribosome pro	790	6	1.2	171	22	ABG00650	Novel human diagno
718	6	1.2	149	22	AAU14202	Human novel protei	791	6	1.2	171	22	ABG16193	Novel human diagno
719	6	1.2	150	22	AAW96200	Purative P. abyssal	792	6	1.2	171	22	ABG23987	Novel human diagno
720	6	1.2	151	21	AAW03418	Rice putative carb	793	6	1.2	172	21	AAW41678	Human ORFX ORF142
721	6	1.2	151	22	AAU32409	Novel human secret	794	6	1.2	173	21	AAW25395	Pinus radiata cell
722	6	1.2	151	22	AAO07440	Human polypeptide	795	6	1.2	173	22	ABG27583	Novel human diagno
723	6	1.2	151	24	ABR40711	Oryza sativa oil t	796	6	1.2	173	22	AAW38749	Novel subtilisin h
724	6	1.2	152	22	AAW79777	Corynebacterium gl	797	6	1.2	173	22	AAU38776	Novel subtilisin h
725	6	1.2	152	22	AAW79945	Corynebacterium gl	798	6	1.2	173	22	AAU38780	Novel subtilisin h
726	6	1.2	152	23	ABW56528	Bm heparan sulfate	799	6	1.2	173	22	AAU38783	Novel subtilisin h
727	6	1.2	152	24	ABP81307	Streptococcus pneu	800	6	1.2	173	22	AAU38799	Novel subtilisin h
728	6	1.2	153	21	AAW10136	Arabidopsis thalia	801	6	1.2	173	22	AAU38802	Novel subtilisin h
729	6	1.2	153	22	ABG20413	Novel human diagno	802	6	1.2	173	22	AAU38807	Novel subtilisin h
730	6	1.2	153	22	ABH12363	Human bone marrow	803	6	1.2	173	22	AAU38808	Novel subtilisin h
731	6	1.2	153	22	AAU30604	Novel human secret	804	6	1.2	173	22	AAU38824	Novel subtilisin h
732	6	1.2	153	22	AAW95559	ERA binding domain	805	6	1.2	173	22	AAU38827	Novel subtilisin h
733	6	1.2	154	23	ABW92249	Herbicideally activ	806	6	1.2	173	22	AAU38845	Novel subtilisin h
734	6	1.2	155	21	AAW28426	Arabidopsis thalia	807	6	1.2	173	22	AAU38846	Novel subtilisin h
735	6	1.2	155	21	AAW42461	Arabidopsis thalia	808	6	1.2	173	22	AAU38848	Novel subtilisin h
736	6	1.2	155	21	AAW71044	Streptococcus pyog	809	6	1.2	173	22	AAU38854	Novel subtilisin h
737	6	1.2	155	22	ABG05294	Novel human diagno	810	6	1.2	173	22	AAU38862	Novel subtilisin h
738	6	1.2	155	22	ABH12375	Human bone marrow	811	6	1.2	173	22	AAU38863	Novel subtilisin h
739	6	1.2	155	22	AAW93696	Human protein sequ	812	6	1.2	173	22	AAU38869	Novel subtilisin h

813	6	1.2	173	22	AAU38870	Novel subtilisin h	886	6	1.2	195	23	ABU52193	Helicobacter pylor
814	6	1.2	173	22	AAU38872	Novel subtilisin h	887	6	1.2	195	23	ABP66916	Human polypeptide
815	6	1.2	174	21	AAU38872	Arabidopsis thalia	888	6	1.2	195	23	ABP47916	Human polypeptide
816	6	1.2	174	22	ABG19365	Novel human diago	889	6	1.2	196	21	AAU39749	Arabidopsis thalia
817	6	1.2	174	22	ABG28588	Novel human diago	890	6	1.2	196	21	AAU39749	Arabidopsis thalia
818	6	1.2	174	23	ABG78123	ITLV, LOR-2, STRI	891	6	1.2	196	21	AAU39749	Arabidopsis thalia
819	6	1.2	174	24	ABG84667	Human SECP-19 prot	892	6	1.2	196	21	AAU39749	Arabidopsis thalia
820	6	1.2	175	17	AAU80793	K3 region reverse	893	6	1.2	196	21	AAU84843	Protein encoded by
821	6	1.2	175	22	AAU85126	C. parvum cryptopa	894	6	1.2	196	21	AAU84843	Soybean pyridoxal
822	6	1.2	175	22	ABJ04062	C. parvum cryptopa	895	6	1.2	196	21	ABG72351	Streptococcus poly
823	6	1.2	177	23	ABG78049	ITLV, LOR-2, STRI	896	6	1.2	198	22	AAU50280	Streptococcus poly
824	6	1.2	177	23	ABP28975	Streptococcus poly	897	6	1.2	198	22	AAU50280	Streptococcus poly
825	6	1.2	178	20	AAU48389	Human prostate can	898	6	1.2	199	14	AAU37520	Amino acid sequenc
826	6	1.2	178	21	AAU27658	Arabidopsis thalia	899	6	1.2	199	14	AAU37520	Arabidopsis thalia
827	6	1.2	178	22	ABG18621	Novel human diago	900	6	1.2	199	17	AAU88661	N. meningitidis B28
828	6	1.2	179	14	AAU50056	ICP34.5 fragment	901	6	1.2	199	19	AAU46806	Aphanoclthe sacrum
829	6	1.2	179	23	ABP29015	Streptococcus poly	902	6	1.2	199	22	AAU35692	Helicobacter pylor
830	6	1.2	180	19	AAU69665	Human NY-ESO-1 pro	903	6	1.2	199	22	AAU35692	Helicobacter pylor
831	6	1.2	180	19	AAU69665	Cancer associated	904	6	1.2	200	23	AAU25617	G protein-coupled
832	6	1.2	180	20	AAU05965	Human cancer antig	905	6	1.2	201	21	AAU10127	Human receptor rei
833	6	1.2	180	21	AAU65990	Human prostate can	906	6	1.2	201	21	AAU10127	Arabidopsis thalia
834	6	1.2	180	21	AAU03154	Human oesophageal	907	6	1.2	201	21	AAU10127	Arabidopsis thalia
835	6	1.2	180	21	AAU70862	Human tumour antig	908	6	1.2	201	21	AAU10127	Arabidopsis thalia
836	6	1.2	180	21	AAU52430	Human tumour antig	909	6	1.2	201	21	AAU52430	Arabidopsis thalia
837	6	1.2	180	22	AAU52430	Novel human diago	910	6	1.2	201	21	AAU52430	Arabidopsis thalia
838	6	1.2	180	22	AAU67164	Amino acid sequenc	911	6	1.2	201	21	AAU64297	Corynebacterium gl
839	6	1.2	180	22	AAU67164	Human NY-ESO-1 pro	912	6	1.2	201	21	AAU64297	Human ORF67. Hom
840	6	1.2	180	22	AAU01535	Human NY-ESO-1 tum	913	6	1.2	203	21	AAU64297	Arabidopsis thalia
841	6	1.2	180	22	AAU01535	Human NY-ESO-1 pro	914	6	1.2	203	21	AAU64297	Arabidopsis thalia
842	6	1.2	180	22	AAU69946	Human protein SEQ	915	6	1.2	203	21	AAU69946	Arabidopsis thalia
843	6	1.2	180	23	AAU05418	Human tumor assoc	916	6	1.2	203	21	AAU05418	Novel human diago
844	6	1.2	180	23	AAU05418	Human tumor assoc	917	6	1.2	203	21	AAU05418	Novel human diago
845	6	1.2	180	24	ABU48210	Human bladder canc	918	6	1.2	204	21	AAU48210	Human DTHP polype
846	6	1.2	180	24	ABU48210	Human NY-ESO-1 pro	919	6	1.2	204	21	AAU48210	Nesleria meningit
847	6	1.2	180	24	ABU56508	Lung cancer-associ	920	6	1.2	204	22	AAU36474	Staphylococcus aur
848	6	1.2	180	24	ABU56508	Lung cancer-associ	921	6	1.2	204	22	AAU36474	Pseudomonas aerugi
849	6	1.2	180	24	ABU74198	Human NY-ESO-1 pro	922	6	1.2	204	22	AAU36474	Staphylococcus aur
850	6	1.2	181	20	AAU37415	Amino acid sequenc	923	6	1.2	204	22	AAU37415	Staphylococcus aur
851	6	1.2	181	22	AAU37415	S. epidermidis ope	924	6	1.2	205	13	AAU37415	Human immune/haema
852	6	1.2	181	22	AAU00538	S. pneumoniae type	925	6	1.2	205	13	AAU37415	Antigen ac-fb. Ei
853	6	1.2	184	21	AAU09243	Arabidopsis thalia	926	6	1.2	206	18	AAU29224	Nitrite hydratase
854	6	1.2	184	21	AAU71039	Streptococcus pyog	927	6	1.2	206	22	AAU71039	Novel human diago
855	6	1.2	184	22	ABG25282	Novel human diago	928	6	1.2	207	14	AAU36389	Urease subunit fro
856	6	1.2	185	23	ABJ11066	Yeast selected int	929	6	1.2	207	21	AAU36389	Arabidopsis thalia
857	6	1.2	185	23	ABG77195	Selected interacti	930	6	1.2	207	22	AAU52771	Arabidopsis thalia
858	6	1.2	187	22	AAU36483	Pseudomonas aerugi	931	6	1.2	207	22	AAU52771	Novel human connec
859	6	1.2	187	24	ABU25838	Aspergillus fumiga	932	6	1.2	207	22	AAU25838	Human ions protein
860	6	1.2	187	24	ABU25838	Aspergillus fumiga	933	6	1.2	207	22	AAU25838	Protonlactactium
861	6	1.2	188	19	AAU62717	Streptococcus pneu	934	6	1.2	207	22	AAU62717	Novel human diago
862	6	1.2	188	24	ABU02799	S. pneumoniae type	935	6	1.2	208	23	ABP41977	Shrimp white spot
863	6	1.2	189	19	AAU02799	S. pneumoniae type	936	6	1.2	208	23	ABP41977	Streptococcus epi
864	6	1.2	189	21	AAU02799	Arabidopsis thalia	937	6	1.2	210	21	AAU02799	Human ovariian anti
865	6	1.2	189	24	ABU02799	S. pneumoniae type	938	6	1.2	210	21	AAU02799	Human cancer assoc
866	6	1.2	190	21	AAU81515	Streptococcus pneu	939	6	1.2	210	21	AAU81515	Human protease and
867	6	1.2	190	22	ABU96208	Human testicular a	940	6	1.2	210	22	ABU96208	Novel human diago
868	6	1.2	190	22	ABU96208	Human reproductive	941	6	1.2	210	22	ABU96208	Novel human diago
869	6	1.2	190	23	ABU55323	Lactococcus lactis	942	6	1.2	210	22	ABU55323	Human novel extrac
870	6	1.2	191	17	AAU04552	Retroviral reverse	943	6	1.2	210	23	ABU55323	Helicobacter pylor
871	6	1.2	191	20	AAU52092	Human retrovirus-5	944	6	1.2	211	17	AAU52092	Human polypeptide
872	6	1.2	191	21	AAU52092	Arabidopsis thalia	945	6	1.2	211	17	AAU52092	N. meningitidis M97
873	6	1.2	191	22	ABG01853	Novel human diago	946	6	1.2	211	21	AAU52092	Human pancreatic c
874	6	1.2	191	22	ABG03072	Novel human diago	947	6	1.2	211	22	AAU52092	Novel human secret
875	6	1.2	191	22	ABG09062	Novel human diago	948	6	1.2	211	22	AAU03694	Group B Streptococ
876	6	1.2	191	22	ABG15048	Novel human diago	949	6	1.2	212	21	AAU03694	Escherichia coli Y
877	6	1.2	192	22	ABG26053	Novel human diago	950	6	1.2	212	21	AAU03694	Arabidopsis thalia
878	6	1.2	193	22	ABU58095	Drosophila melanog	951	6	1.2	212	21	AAU58095	Arabidopsis thalia
879	6	1.2	194	22	AAU19887	Human novel extrac	952	6	1.2	213	21	AAU19887	LI UB4-domain prot
880	6	1.2	194	23	ABP48107	Human polypeptide	953	6	1.2	213	21	ABP48107	Human ORF ORF308
881	6	1.2	195	22	ABU10329	Human cDNA SEQ ID	954	6	1.2	213	21	ABU10329	Human Rab4b protei
882	6	1.2	195	22	AAU18492	Human endocrine po	955	6	1.2	213	21	AAU18492	drab4 amino acid s
883	6	1.2	195	22	AAU36444	Human polypeptide	956	6	1.2	213	24	AAU36444	rab4b amino acid
884	6	1.2	195	22	AAU19696	Human novel extrac	957	6	1.2	214	21	AAU19696	S. pneumoniae type
885	6	1.2	195	22	AAU21700	Novel human neopla	958	6	1.2	214	21	AAU21700	Arabidopsis thalia

959	6	1.2	214	22	AAU17093	Novel signal trans
960	6	1.2	214	23	AAO18339	Mature humanised m
961	6	1.2	215	21	AAO19526	Arabidopsis thalia
962	6	1.2	216	21	AAO57236	Arabidopsis thalia
963	6	1.2	216	21	AAO61182	Arabidopsis thalia
964	6	1.2	216	22	ABE5053	Drosophila melanog
965	6	1.2	217	21	AAV1042	Streptococcus pyog
966	6	1.2	217	22	ABE60591	Drosophila melanog
967	6	1.2	217	22	AAU41274	Propionibacterium
968	6	1.2	217	22	AAU42502	Propionibacterium
969	6	1.2	217	23	ABP30013	Streptococcus poly
970	6	1.2	218	21	ABR1633	Pinus radiata alph
971	6	1.2	218	22	ABR70933	Drosophila melanog
972	6	1.2	219	21	AAV6981	Quail H1 histone p
973	6	1.2	219	23	ABP29022	Streptococcus poly
974	6	1.2	222	20	AAV36900	Chlamydia trachoma
975	6	1.2	222	22	ABG08919	Novel human diagno
976	6	1.2	222	22	AAU17547	Novel signal trans
977	6	1.2	223	22	AAU30182	Novel human secret
978	6	1.2	224	22	AAW62229	Sma-4 protein C-te
979	6	1.2	224	21	AAW87590	Aminoamidase pre
980	6	1.2	224	22	ABG01822	Novel human diagno
981	6	1.2	224	23	ABR81607	A. thaliana glucos
982	6	1.2	225	22	AAU35452	Haemophilus influe
983	6	1.2	225	22	ABG02862	Novel human diagno
984	6	1.2	225	22	ABG03051	Novel human diagno
985	6	1.2	225	22	AAU27631	Human protein AFP2
986	6	1.2	225	22	AAU17137	Novel signal trans
987	6	1.2	226	21	AAO20903	Arabidopsis thalia
988	6	1.2	226	21	AAO37582	Arabidopsis thalia
989	6	1.2	226	22	ABE65906	Drosophila melanog
990	6	1.2	226	22	AAAB3964	Partial amino acid
991	6	1.2	226	23	ABR81606	A. thaliana glucos
992	6	1.2	227	22	ABG16125	Novel human diagno
993	6	1.2	228	23	ABU51752	Helicobacter pylori
994	6	1.2	229	20	AAV29178	Amino acid sequenc
995	6	1.2	229	21	AAO41580	Arabidopsis thalia
996	6	1.2	229	23	ABG66716	Human novel polype
997	6	1.2	230	21	AAO29871	Arabidopsis thalia
998	6	1.2	233	16	AAO67918	(1-3)-beta-D-gluc
999	6	1.2	233	23	ABE53558	Bifidobacterium 10
1000	6	1.2	233	23	ABE5603	Bifidobacterium 10

## ALIGNMENTS

## RESULT 1

AAO67668 standard; Protein; 502 AA.

XX	AAO67668	
AC	AAO67668	
DT	11-JUN-2001 (first entry)	
XX	Amino acid sequence of a BASB109 polypeptide.	
DE	BASB109; bacterial infection; vaccine; genetic immunisation.	
XX	Moraxella catarrhalis.	
OS	MO200119996-A1.	
PN	22-MAR-2001.	
PD	14-SEP-2000; 2000WO-EPO9035.	
PF	14-SEP-1999; 99GB-0021691.	
PR	(SMIK) SMITHKLINE BEECHAM BIOLOGICALS.	
XX	Thomard J;	
XX		

DR	WPI; 2001-257883/26.
DR	N-PSDB; AAF55658.
XX	
PT	Novel BASB109 polypeptides of Moraxella catarrhalis useful for
PT	diagnosis, prophylactic and therapeutic purposes against microbial
PT	diseases, preferably bacterial infections
XX	
PS	Claim 3; Page 65; 93pp; English.
XX	
CC	The present sequence represents a BASB109 polypeptide of Moraxella
CC	catarrhalis. BASB109 polypeptides and polynucleotides are useful for
CC	treating bacterial infections, and as research reagents and materials
CC	for the treatment of and diagnosis of diseases, particularly human
CC	diseases. They are useful for inducing an immune response in an
CC	individual, and to assess the binding of small molecule substrates and
CC	ligands in, for e.g. cells, cell-free preparations, chemical libraries,
CC	and natural product mixtures. BASB109 polynucleotides are useful for
CC	therapeutic or prophylactic purposes, in particular genetic immunisation
CC	and in diagnosis of the stage and type of infection.
XX	
SO	Sequence 502 AA;
Query Match	100.0%; Score 502; DB 22; Length 502;
Best Local Similarity	100.0%; Pred. No. 0;
Matches 502; Conservative	0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MSKPTIKTTLCALSMALSGCSNOADKAQPKSTVAAAKTANADNAASOEHOGEHP	60
DB	1	MSKPTIKTTLCALSMALSGCSNOADKAQPKSTVAAAKTANADNAASOEHOGEHP	60
QY	61	VIDAIYTHAPEVPPVDRDHPAKVVKMETEKVKMLADVEYQFTFGGOVPGQMRVR	120
DB	61	VIDAIYTHAPEVPPVDRDHPAKVVKMETEKVKMLADVEYQFTFGGOVPGQMRVR	120
QY	121	EGDTIVQSNRPHDSMPHNVDPHATGPGGAASFTAPGHTSTFSFALQPLVYHC	180
DB	121	EGDTIVQSNRPHDSMPHNVDPHATGPGGAASFTAPGHTSTFSFALQPLVYHC	180
QY	181	AVAPVGMHANGVYGLILVEPKEGLPKVDKEYVMQGDYTYGKYGEGLQPPDMEKAR	240
DB	181	AVAPVGMHANGVYGLILVEPKEGLPKVDKEYVMQGDYTYGKYGEGLQPPDMEKAR	240
QY	241	EDAEYVVFNGSVGALTGENALAKAVGETVRLFVNGGPNLTSSFVIGEIFDKVHEGK	300
DB	241	EDAEYVVFNGSVGALTGENALAKAVGETVRLFVNGGPNLTSSFVIGEIFDKVHEGK	300
QY	301	GENHNIQTLLIPAGGAITEFKVDVPGDVLYVDHAFRAFNGALGILVEGEENHEIYS	360
DB	301	GENHNIQTLLIPAGGAITEFKVDVPGDVLYVDHAFRAFNGALGILVEGEENHEIYS	360
QY	361	HKOTDAVYLPFGAPQAIIDQZAPKTPAPANLQEOIKAGATYDSNCAACHOPDGKVPNA	420
DB	361	HKOTDAVYLPFGAPQAIIDQZAPKTPAPANLQEOIKAGATYDSNCAACHOPDGKVPNA	420
QY	421	FPEPLANSVYLNADHAPASIVANGLSKITVNGNOYESVMPALISDOQIANVITYTLNS	480
DB	421	FPEPLANSVYLNADHAPASIVANGLSKITVNGNOYESVMPALISDOQIANVITYTLNS	480
QY	481	FGKGGQLSADVAKAKTKTPN	502
DB	481	FGKGGQLSADVAKAKTKTPN	502

## RESULT 2

AAO67669 standard; Protein; 502 AA.

XX	AAO67669	
AC	AAO67669	
DT	11-JUN-2001 (first entry)	
XX	Amino acid sequence of a BASB109 polypeptide.	
DE		
XX		

KM BASB109; bacterial infection; vaccine; genetic immunisation.  
 XX Moraxella catarrhalis.  
 OS WO200119996-A1.  
 XX  
 XX PD 22-MAR-2001.  
 XX PF 14-SEP-2000; 2000WO-BE09035.  
 XX PR 14-SEP-1999; 99GB-0021691.  
 XX (SMK ) SMITHKLINE BEECHAM BIOLOGICALS.  
 PA Thonnard J;  
 XX PI  
 XX DR WPI: 2001-257883/26.  
 DR N-PSDB; AAF55657.  
 XX  
 PT Novel BASB109 polypeptides of Moraxella catarrhalis useful for  
 PT diagnostic, prophylactic and therapeutic purposes against microbial  
 PT diseases, preferably bacterial infections  
 XX  
 PS Claim 3; Page 66; 93pp; English.  
 CC The present sequence represents a BASB109 polypeptide of Moraxella  
 CC catarrhalis. BASB109 polypeptides and polynucleotides are useful for  
 CC treating bacterial infections, and as research reagents and materials  
 CC for the treatment of and diagnosis of diseases, particularly human  
 CC diseases. They are useful for inducing an immune response in an  
 CC individual, and to assess the binding of small molecule substrates and  
 CC ligands in, for e.g. cells, cell-free preparations, chemical libraries,  
 CC and natural product mixtures. BASB109 polynucleotides are useful for  
 CC therapeutic or prophylactic purposes, in particular genetic immunisation  
 CC and in diagnosis of the stage and type of infection.  
 XX  
 SQ Sequence 502 AA;  
 Query Match 100.0%; Score 502; DB 22; Length 502;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 502; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MSKPTLITKTLTICALSALMLSCGSGNADRAQPKSSTVDAAKTANADNAAQEHGELP 60  
 DB 1 MSKPTLITKTLTICALSALMLSCGSGNADRAQPKSSTVDAAKTANADNAAQEHGELP 60  
 QY 61 VIDAIVTAHPVPPVDRDHPAKVYVYKMEVEMKLAGVCEYQFTFGQVPGQIRYR 120  
 DB 61 VIDAIVTAHPVPPVDRDHPAKVYVYKMEVEMKLAGVCEYQFTFGQVPGQIRYR 120  
 QY 121 EGDTEVQSNHPDSKPHNVDPFAATGPGGGAESAFTAPGHTSTPSFKALQGLVYHC 180  
 DB 121 EGDTEVQSNHPDSKPHNVDPFAATGPGGGAESAFTAPGHTSTPSFKALQGLVYHC 180  
 QY 181 AVAPVGMHIANGYGLIIVEPEKGLPKVKEYVVMQGDFTYKKGVEGGLQPFDMKAR 240  
 DB 181 AVAPVGMHIANGYGLIIVEPEKGLPKVKEYVVMQGDFTYKKGVEGGLQPFDMKAR 240  
 QY 241 EDAEYVVFNGSVGALTGEMALAKKVGSEYRLLFVNGNGPMLTSSFHVIGELFDVHVEGSK 300  
 DB 241 EDAEYVVFNGSVGALTGEMALAKKVGSEYRLLFVNGNGPMLTSSFHVIGELFDVHVEGSK 300  
 QY 301 GENNHIQTLTIPAGGAITEFKVDVGDVYLVDAHFAFNKALGILVGEENHEIYS 360  
 DB 301 GENNHIQTLTIPAGGAITEFKVDVGDVYLVDAHFAFNKALGILVGEENHEIYS 360  
 QY 361 HKQTDVAVLPEGAPOAIDTQEAKPAPANLQEOITAGATYDSNCAQHPDGKGVNA 420  
 DB 361 HKQTDVAVLPEGAPOAIDTQEAKPAPANLQEOITAGATYDSNCAQHPDGKGVNA 420  
 QY 421 FPLPLNSDVLNADHAAASIVANGLSGKTTVNGNYESVMPAIALSDOOIANYITTLNS 480  
 DB 421 FPLPLNSDVLNADHAAASIVANGLSGKTTVNGNYESVMPAIALSDOOIANYITTLNS 480

QY 481 FGNKGGQLSADDVAKAKTKTPN 502  
 DB 481 FGNKGGQLSADDVAKAKTKTPN 502  
 RESULT 3  
 AAO17561  
 ID AAO17561 standard; Protein; 197 AA.  
 XX  
 AC AAO17561;  
 XX  
 DT 19-UTL-2002 (first entry)  
 DE  
 DE M catarrhalis MCA100170 protein SEQ ID NO: 2.  
 XX  
 XX Moraxella; vaccine; respiratory tract infection; antiinflammatory;  
 KM auditory; antibacterial; otitis media; sinusitis; pneumonia.  
 OS Moraxella catarrhalis.  
 XX  
 PN WO200218595-A2.  
 XX  
 PD 07-MAR-2002.  
 XX  
 XX PE 28-AUG-2001; 2001WO-CA01221.  
 XX  
 PR 28-AUG-2000; 2000US-228294P.  
 PR 28-AUG-2000; 2000US-228295P.  
 PR 28-AUG-2000; 2000US-228296P.  
 PR 29-AUG-2000; 2000US-228438P.  
 PR 29-AUG-2000; 2000US-228439P.  
 PR 29-AUG-2000; 2000US-228440P.  
 PR 29-AUG-2000; 2000US-228441P.  
 PR 29-AUG-2000; 2000US-228442P.  
 PR 29-AUG-2000; 2000US-228443P.  
 PR 29-AUG-2000; 2000US-228511P.  
 PR 29-AUG-2000; 2000US-228512P.  
 PR 29-AUG-2000; 2000US-228742P.  
 PR 29-AUG-2000; 2000US-228773P.  
 PR 01-SEP-2000; 2000US-229465P.  
 PR 01-SEP-2000; 2000US-229474P.  
 PR 01-SEP-2000; 2000US-229475P.  
 PR 01-SEP-2000; 2000US-229478P.  
 PR 05-SEP-2000; 2000US-229740P.  
 PR 05-SEP-2000; 2000US-229803P.  
 PR 05-SEP-2000; 2000US-229804P.  
 PR 05-SEP-2000; 2000US-229805P.  
 PR 05-SEP-2000; 2000US-229806P.  
 PR 05-SEP-2000; 2000US-229809P.  
 PR 05-SEP-2000; 2000US-229811P.  
 PR 06-SEP-2000; 2000US-230214P.  
 PR 06-SEP-2000; 2000US-230250P.  
 PR 06-SEP-2000; 2000US-230252P.  
 PA (AVET ) AVENTIS PASTEUR LTD.  
 XX  
 PI Loesmore S, Wang J, Bradley B, Ochs M, Yang Y;  
 DR WPI: 2002-401721/43.  
 DR N-PSDB; AAL46493.  
 XX  
 PT Moraxella polypeptide and polynucleotides useful as vaccine for  
 PT immunizing a host e.g. humans against disease e.g. otitis media,  
 PT pneumonia, caused by infection of the bacteria  
 XX  
 PS Claim 28; Fig 1; 277pp; English.  
 CC The present invention provides the protein and coding sequences of  
 CC proteins from Moraxella catarrhalis. These can be used to produce  
 CC vaccines which protect against M. catarrhalis infection, which can cause  
 CC otitis media, respiratory infection, sinusitis, and pneumonia. The  
 CC present sequence is a protein of the invention.

XX Sequence 197 AA;  
 SQ Query Match 39.2%; Score 197; DB 23; Length 197;  
 Best Local Similarity 100.0%; Pred. No. 7.4e-183;  
 Matches 197; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSKPTLTKTLTICALSALMLSGCSNOADKKAQPKSSTVDAAKTANADNMAOEHQGLP 60  
 DB 1 MSKPTLTKTLTICALSALMLSGCSNOADKKAQPKSSTVDAAKTANADNMAOEHQGLP 60  
 QY 61 VIDAIVTTHAPPPPPVDRDHPAKVYVVKETVEKVMRLADGVEYQFTFGQVPCQMIRV 120  
 DB 61 VIDAIVTTHAPPPPPVDRDHPAKVYVVKETVEKVMRLADGVEYQFTFGQVPCQMIRV 120  
 QY 121 EGDITIEVOFSNHPDSKMPHNVDFAATGPGGAEASFTAPGHTSTSPKALQPLVYVHC 180  
 DB 121 EGDITIEVOFSNHPDSKMPHNVDFAATGPGGAEASFTAPGHTSTSPKALQPLVYVHC 180  
 QY 181 AVAPVGMHIANGMVGLI 197  
 DB 181 AVAPVGMHIANGMVGLI 197

RESULT 4  
 ABP77627 ID ABP77627 standard; Protein; 392 AA.  
 XX ABP77627;  
 AC  
 XX 07-MAR-2003 (first entry)  
 DT  
 XX N. gonorrhoeae amino acid sequence SEQ ID 1784.  
 DE  
 XX Antibacterial; infection; vaccine; gene therapy.  
 KM  
 XX Neisseria gonorrhoeae.  
 OS  
 XX WO200279243-A2.  
 PN  
 XX 10-OCT-2002.  
 PD  
 XX 12-FEB-2002; 2002WO-IB02069.  
 PF  
 XX 12-FEB-2001; 2001GB-0003424.  
 PR  
 XX (CHIR-) CHIRON SPA.  
 PA  
 XX Fontana MR, Piazza M, Massignani V, Monaci E;  
 PI WPI; 2003-058415/05.  
 DR N-PSDB; AB238597.  
 DR  
 XX New protein from Neisseria gonorrhoeae, useful for the manufacture of a  
 PT medicament for treating or preventing N. gonorrhoeae infection -  
 PS Disclosure; Page 316; 815pp; English.  
 XX The present invention relates to proteins from Neisseria gonorrhoeae.  
 CC Also disclosed are the nucleic acid molecules encoding the proteins and  
 CC antibodies that specifically bind to the proteins. The composition  
 CC comprising the protein, nucleic acid or antibody is useful for the  
 CC manufacture of a medicament for treating or preventing N. gonorrhoeae  
 CC infection, this may be in the form of a vaccine or gene therapy.  
 CC Sequences given in records ABP76736-ABP81046 represent nucleic acid  
 CC molecules of the invention.  
 XX  
 SQ Sequence 392 AA;  
 Query Match 6.8%; Score 34; DB 24; Length 392;  
 Best Local Similarity 100.0%; Pred. No. 3.4e-24;  
 Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 178 YHCAPAVPGMHIANGMVGLIVPEKGLPKVDKE 211  
 DB 173 YHCAPAVPGMHIANGMVGLIVPEKGLPKVDKE 206

RESULT 5  
 AAY05975 ID AAY05975 standard; Peptide; 10 AA.  
 XX AAY05975;  
 AC  
 XX 16-AUG-1999 (first entry)  
 DT  
 XX Human cancer antigen NY ESO-1/CAG-3 ORF1 cancer peptide.  
 DE  
 XX NY ESO-1/CAG-3 gene; CAG-3 gene; cancer peptide; antigen; human;  
 KM leukemia; non-Hodgkins lymphoma; Hodgkins lymphoma; lung cancer;  
 KM metastasis; melanoma; adenocarcinoma; thymoma; colon cancer;  
 KM uterine cancer; breast cancer; prostate cancer; ovarian cancer;  
 KM cervical cancer; bladder cancer; kidney cancer; pancreatic cancer;  
 KM liver cancer; sarcoma; tumour; diagnosis; immunotherapy; therapy;  
 KM vaccine.  
 XX  
 XX Homo sapiens.  
 OS Synthetic.  
 OS  
 XX WO9918206-A2.  
 PN  
 XX 15-APR-1999.  
 PD  
 XX 21-SEP-1998; 98WO-US19609.  
 PF  
 XX 08-OCT-1997; 97US-0061428.  
 PR  
 XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
 PA  
 XX Rosenberg SA, Wang RF;  
 PI WPI; 1999-277270/23.  
 DR  
 XX Cancer antigen NY ESO1/CAG-3  
 PT  
 XX Claim 29; Page 10; 88pp; English.  
 PS  
 XX The present sequence represents a cancer peptide that is based on  
 CC amino acid residues 55-62 of human ESO-1/CAG-3 (or CAG-3) ORF1  
 CC (see AAY05965), a new and potent tumour antigen capable of eliciting  
 CC an antigen specific immune response by T cells. Cancer peptides  
 CC derived from CAG-3 ORF1, CAG-3 ORF2 (see AAY05966), portions of them  
 CC and their variants (see AAY05967-87), are useful as cancer vaccines  
 CC that protect against cancer. The invention provides: vectors and  
 CC host cells (also useful as vaccines); a method of diagnosis of  
 CC cancer or precancer; a transgenic animal; antisense oligonucleotides  
 CC that inhibit expression of the cancer peptide or tumour antigen;  
 CC antibodies reacting with a CAG-3 cancer peptide, useful in  
 CC diagnostic and detection assays; and methods for preventing or  
 CC inhibiting cancer by administering a cancer peptide, with or without  
 CC an HLA molecule. The cancer peptides form part of, or are derived  
 CC from, cancers such as primary or metastatic melanoma, thymoma,  
 CC lymphoma, sarcoma, lung cancer, liver cancer, leukemia, uterine  
 CC cancer, cervical cancer, bladder cancer, kidney cancer and  
 CC adenocarcinomas such as breast, prostate, ovarian, pancreatic and  
 CC thyroid cancers. Melanoma is treated by inducing cancer-specific T  
 CC cells in vitro for subsequent return to a patient.  
 XX  
 SQ Sequence 10 AA;  
 Query Match 1.6%; Score 8; DB 20; Length 10;  
 Best Local Similarity 100.0%; Pred. No. 2.6;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



Db 1 ATPGCGA 8

RESULT 6  
ABG18127  
ID ABG18127 standard; Protein; 415 AA.  
XX  
XX  
AC ABG18127;  
XX  
DT 18-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #18118.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
OS Homo sapiens.  
XX  
FN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
XX  
PR 23-AUG-2000; 2000US-0649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Dmanac RT, Liu C, Tang YT;  
XX  
DR WPI; 2001-639362/73.  
XX  
DR N-PSDB; AAS82314.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX  
XX  
PS Claim 20; SEQ ID No 48486; 103pp; English.  
XX  
XX  
CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostic, forensic, gene mapping, identification of mutations in  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG0010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pat\_sequences.  
XX  
XX  
SQ Sequence 415 AA;  
Query Match 1.6%; Score 8; DB 22; Length 415;  
Best Local Similarity 100.0%; Pred. No. 65;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 434 HARRASIV 441  
Db 250 HARRASIV 257

RESULT 7  
AAV74923  
ID AAV74923 standard; Protein; 497 AA.  
XX  
XX  
AC AAV74923;  
XX  
DT 21-MAR-2000 (first entry)  
XX  
DE Neisseria gonorrhoeae ORF 402 protein sequence SEQ ID NO:11320.  
XX  
KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;  
KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;  
KW antibacterial; gene therapy.  
XX  
OS Neisseria gonorrhoeae.  
XX  
FN WO9957280-A2.  
XX  
PD 11-NOV-1999.  
XX  
PF 30-APR-1999; 99WO-US09346.  
XX  
PR 01-MAY-1998; 98US-0083758.  
XX  
PR 31-JUL-1998; 98US-0094869.  
XX  
PR 02-SEP-1998; 98US-0098994.  
XX  
PR 02-SEP-1998; 98US-0099062.  
XX  
PR 09-OCT-1998; 98US-0103749.  
XX  
PR 09-OCT-1998; 98US-0103794.  
XX  
PR 09-OCT-1998; 98US-0103796.  
XX  
PR 25-FEB-1999; 99US-0121528.  
XX  
PA (CHIR ) CHIRON CORP.  
PA (GENO-) INST GENOMIC RES.  
XX  
XX  
PI Frazer C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;  
PI Petersen J, Piza M, Rappoli R, Ratti G, Scalato E, Scarselli M;  
PI Tettelin H, Venter JC;  
XX  
XX  
DR WPI; 2000-062150/05.  
XX  
DR N-PSDB; AAZ53685.  
XX  
PT Novel Neisserial polypeptides predicted to be useful antigens for  
PT vaccines and diagnostics -  
XX  
XX  
PS Claim 2; Page 724; 1453pp; English.  
XX  
XX  
CC AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAV74253 to AAV75941  
CC represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides  
CC and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ5473 represent  
CC PCR primers used in the exemplification of the present invention. The  
CC polypeptides, the polynucleotides, antibodies and compositions of  
CC the invention can be used as vaccines, as diagnostic reagents, and as  
CC immunogenic compositions. The polypeptides can be used in the  
CC manufacture of medicaments for treating or preventing infection due to  
CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the  
CC presence of Neisseria bacteria, or to raise antibodies. They may also  
CC be used to screen for agonists or antagonists, which may themselves  
CC have use as antibacterial agents. The polynucleotides of the invention  
CC may also be used in gene therapy protocols.  
XX  
XX  
SQ Sequence 497 AA;  
Query Match 1.6%; Score 8; DB 21; Length 497;  
Best Local Similarity 100.0%; Pred. No. 76;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 35 SSTVDAAA 42  
Db 454 SSTVDAAA 461

RESULT 8

AAV74924  
 ID AAV74924 standard; Protein; 497 AA.  
 XX  
 AC AAV74924;  
 XX  
 DT 21-MAR-2000 (first entry)  
 XX  
 DE Neisseria meningitidis ORF 402 protein sequence SEQ ID NO:1322.  
 XX  
 KM Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;  
 KM antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;  
 KM antibacterial; gene therapy.  
 XX  
 OS Neisseria meningitidis.  
 XX  
 PN WO957280-A2.  
 XX  
 PD 11-NOV-1999.  
 XX  
 PF 30-APR-1999; 99WO-US09346.  
 XX  
 PR 01-MAY-1998; 98US-0083758.  
 PR 31-JUL-1998; 98US-0094869.  
 PR 02-SEP-1998; 98US-0098994.  
 PR 02-SEP-1998; 98US-0099062.  
 PR 09-OCT-1998; 98US-0103749.  
 PR 09-OCT-1998; 98US-0103794.  
 PR 09-OCT-1998; 98US-0103796.  
 PR 25-FEB-1999; 99US-0121528.  
 XX  
 PA (CHIR ) CHIRON CORP.  
 PA (GENO-) INST GENOMIC RES.  
 XX  
 PI Fraser C, Galeotti C, Grandi G, Hickey E, Masiagnani V, Mora M;  
 PI Petersen J, Pizsa M, Rappuoli R, Ratti G, Scalato E, Scarselli M;  
 PI Tettelin H, Venter JC.  
 XX  
 DR WPI; 2000-062150/05.  
 DR N-PSDB; AA253686.  
 XX  
 PT Novel Neisserial polypeptides predicted to be useful antigens for  
 PT vaccines and diagnostics -  
 XX  
 PS Claim 2; Page 725; 1453pp; English.  
 XX  
 CC AA253015 to AA254536, AA254577 to AA254615, and AAV74253 to AAV75941  
 CC represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides  
 CC and polypeptides. AA254537 to AA254576 and AA254616 to AA25473 represent  
 CC PCR primers used in the exemplification of the present invention. The  
 CC polypeptides, the polynucleotides, antibodies and compositions of  
 CC the invention can be used as vaccines, as diagnostic reagents, and as  
 CC immunogenic compositions. The polypeptides can be used in the  
 CC manufacture of medicaments for treating or preventing infection due to  
 CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the  
 CC presence of Neisseria bacteria, or to raise antibodies. They may also  
 CC be used to screen for agonists or antagonists, which may themselves  
 CC have use as antibacterial agents. The polynucleotides of the invention  
 CC may also be used in gene therapy protocols.  
 XX  
 SQ Sequence 497 AA;  
 XX  
 Query Match 1.6%; Score 8; DB 21; Length 497;  
 Best Local Similarity 100.0%; Pred. No. 76;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 35 SSTVDAAA 42  
 |||||  
 Db 454 SSTVDAAA 461

XX  
 AC AAV74925;  
 XX  
 DT 21-MAR-2000 (first entry)  
 XX  
 DE Neisseria meningitidis ORF 402 protein sequence SEQ ID NO:1324.  
 XX  
 KM Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;  
 KM antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;  
 KM antibacterial; gene therapy.  
 XX  
 OS Neisseria meningitidis.  
 XX  
 PN WO957280-A2.  
 XX  
 PD 11-NOV-1999.  
 XX  
 PF 30-APR-1999; 99WO-US09346.  
 XX  
 PR 01-MAY-1998; 98US-0083758.  
 PR 31-JUL-1998; 98US-0094869.  
 PR 02-SEP-1998; 98US-0098994.  
 PR 02-SEP-1998; 98US-0099062.  
 PR 09-OCT-1998; 98US-0103749.  
 PR 09-OCT-1998; 98US-0103794.  
 PR 09-OCT-1998; 98US-0103796.  
 PR 25-FEB-1999; 99US-0121528.  
 XX  
 PA (CHIR ) CHIRON CORP.  
 PA (GENO-) INST GENOMIC RES.  
 XX  
 PI Fraser C, Galeotti C, Grandi G, Hickey E, Masiagnani V, Mora M;  
 PI Petersen J, Pizsa M, Rappuoli R, Ratti G, Scalato E, Scarselli M;  
 PI Tettelin H, Venter JC.  
 XX  
 DR WPI; 2000-062150/05.  
 DR N-PSDB; AA253687.  
 XX  
 PT Novel Neisserial polypeptides predicted to be useful antigens for  
 PT vaccines and diagnostics -  
 XX  
 PS Claim 2; Page 726-727; 1453pp; English.  
 XX  
 CC AA253015 to AA254536, AA254577 to AA254615, and AAV74253 to AAV75941  
 CC represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides  
 CC and polypeptides. AA254537 to AA254576 and AA254616 to AA25473 represent  
 CC PCR primers used in the exemplification of the present invention. The  
 CC polypeptides, the polynucleotides, antibodies and compositions of  
 CC the invention can be used as vaccines, as diagnostic reagents, and as  
 CC immunogenic compositions. The polypeptides can be used in the  
 CC manufacture of medicaments for treating or preventing infection due to  
 CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the  
 CC presence of Neisseria bacteria, or to raise antibodies. They may also  
 CC be used to screen for agonists or antagonists, which may themselves  
 CC have use as antibacterial agents. The polynucleotides of the invention  
 CC may also be used in gene therapy protocols.  
 XX  
 SQ Sequence 497 AA;  
 XX  
 Query Match 1.6%; Score 8; DB 21; Length 497;  
 Best Local Similarity 100.0%; Pred. No. 76;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 35 SSTVDAAA 42  
 |||||  
 Db 454 SSTVDAAA 461

RESULT 9  
 AAV74925  
 ID AAV74925 standard; Protein; 497 AA.

RESULT 10  
 ABP80199  
 ID ABP80199 standard; Protein; 508 AA.  
 XX  
 AC ABP80199;

```

XX      07-MAR-2003   (first entry)
DT
XX
XX      N. gonorrhoeae amino acid sequence SEQ ID 6928.
DE
XX      Antibacterial; infection; vaccine; gene therapy.
KW
XX      Neisseria gonorrhoeae.
OS
XX      WO200279243-A2.
PN
XX      10-OCT-2002.
PD
XX      12-FEB-2002; 2002WO-IB02069.
PE
XX      12-FEB-2001; 2001GB-0003424.
PR
XX      (CHIR-) CHIRON SPA.
PA
XX      Fontana MR, Pizza M, Maignani V, Monaci E;
PI      WPI; 2003-058415/05.
XX      DR
XX      N-PSDB; ABZ41169.
DR
XX      New protein from Neisseria gonorrhoeae, useful for the manufacture of a
PT      medicament for treating or preventing N. gonorrhoeae infection -
PT
XX      Disclosure; Page 690; 815pp; English.
PS
XX      The present invention relates to proteins from Neisseria gonorrhoeae.
CC      Also disclosed are the nucleic acid molecules encoding the proteins and
CC      antibodies that specifically bind to the proteins. The composition
CC      comprising the protein, nucleic acid or antibody is useful for the
CC      manufacture of a medicament for treating or preventing N. gonorrhoeae
CC      infection, this may be in the form of a vaccine or gene therapy.
CC      Sequences given in records ABP76736-ABP81046 represent nucleic acid
CC      molecules of the invention.
SQ
XX      Sequence     508 AA;

Query Match          1.6%; Score 8; DB 24; Length 508;
Best Local Similarity 100.0%; Pred. No. 77;
Matches      8; Conservative    0; Mismatches    0; Indels    0; Gaps    0,
QY      35 SSTVDAAA 42
        |||||
Db       465 SSTVDAAA 472

RESULT 11
AAID ID AAY93290 standard; Protein; 697 AA.
AC AAY93290;
AD
DT 04-SEP-2000 (first entry)
DE Amino acid sequence of a polypeptide of a Neisseria pathogenic strain.
DS
KW Pathogenic strain; Neisseria; vaccine; Neisseria infection.
OS Neisseria meningitidis.
PN WO200026375-A2.
PD
PE 11-MAY-2000.
PF 28-OCT-1999; 99WO-FR02643.
PG
PH 30-OCT-1998; 98FR-0013693.
PI
PP (INMR ) PASTEUR MERIEUX SERUMS & VACCINS SA.
PR (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
PT

```

XX Aujame L, Bouchardon A, Renaud-Mongenie G, Rokbi B, Nassif X;  
PI Tinsley C, Perrin A;  
XX  
DR WPI; 2000-365622/31.  
DR N-PSDB; AAA15321.  
XX  
PT New polypeptide specific for pathogenic *Neisseria* useful in therapeutic  
PT or preventative vaccines and for diagnosis -  
XX  
PS Claim 5; Page 128-131; 187pp; French.  
XX  
XX The present sequence represents a protein that is specific for pathogenic  
CC strains of *Neisseria*. The polynucleotides, polypeptides, or their  
CC antigenic fragments, are used in vaccines to treat or protect against  
CC *Neisseria* infections, particularly by *N. meningitidis*. The  
CC polynucleotide sequence is also used for recombinant production of  
CC the polypeptide and to produce attenuated *Neisseria* strains that  
CC overexpress it, or express it in a non-toxic mutant form.  
SQ Sequence 697 AA;  
  
OY 35 SSTVDAAA 42  
Db 654 SSTVDAAA 661  
  
RESULT 12  
AA93303  
ID AAY93303 standard; Protein; 697 AA.  
XX  
XX AAY93303;  
DT 04-SEP-2000 (first entry)  
XX  
XX Amino acid sequence of a polypeptide of a *Neisseria* pathogenic strain.  
DE Pathogenic strain; *Neisseria*; vaccine; *Neisseria* infection.  
XX  
XX *Neisseria* gonorrhoeae.  
OS  
XX WO200026375-A2.  
XX  
XX 11-MAY-2000.  
XX  
XX 28-OCT-1999; 99WO-FR02643.  
XX  
XX 30-OCT-1998; 98FR-0013693.  
XX  
XX (INMR ) PASTEUR MERIEUX SERUMS & VACCINS SA.  
PA (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.  
XX  
XX Aujame L, Bouchardon A, Renaud-Mongenie G, Rokbi B, Nassif X;  
PI Tinsley C, Perrin A;  
XX  
XX WPI; 2000-365622/31.  
DR N-PSDB; AAA15333.  
XX  
XX New polypeptide specific for pathogenic *Neisseria* useful in therapeutic  
PT or preventative vaccines and for diagnosis -  
XX  
PS Claim 6; Page 173-175; 187pp; French.  
XX  
XX The present sequence represents a protein that is specific for pathogenic  
CC strains of *Neisseria*. The polynucleotides, polypeptides, or their  
CC antigenic fragments, are used in vaccines to treat or protect against  
CC *Neisseria* infections, particularly by *N. meningitidis*. The  
CC polynucleotide sequence is also used for recombinant production of  
CC the polypeptide and to produce attenuated *Neisseria* strains that

CC overexpress it, or express it in a non-toxic mutant form.  
XX  
SQ Sequence 697 AA;

Query Match 1.6%; Score 8; DB 21; Length 697;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 35 SSTVDAAA 42  
|||  
Db 654 SSTVDAAA 661

RESULT 13  
ABP65660  
ID ABP65660 standard; Protein; 995 AA.

AC ABP65660;

DT 19-NOV-2002 (first entry)

DE Bifidobacterium longum NCC2705 ORF amino acid sequence SEQ ID NO:404.

XX Bifidobacterium longum NCC2705; Bifidobacterium; bacterial;  
KW anti-diarrhetic; antibacterial; inhibitor of Salmonella; detection;  
KM identification; lactic acid bacterium; diarrhoea; pathogenic bacteria;  
XX rotavirus; food composition; pharmaceutical composition.

OS Bifidobacterium longum.

PN EPI227152-A1.

PD 31-JUL-2002.

PF 30-JAN-2001; 2001EP-0102050.

PR 30-JAN-2001; 2001EP-0102050.

XX (NEST) SOC PROD NESTLE SA.

XX WPI; 2002-668397/72.

PT Novel polynucleotide comprising Bifidobacterium genome sequence useful  
as a probe or primer for detecting and/or identifying Bifidobacterium  
longum in a biological sample -

PS Claim 3; SEQ ID 404; 80bp; English.

XX The present invention describes a polynucleotide (I) comprising a  
CC sequence of a Bifidobacterium genome selected from the nucleotide  
CC sequences given in AB081842 and AB081843, or a sequence exhibiting at  
CC least 90% identity or which hybridizes with the sequences given in  
CC AB081842 and AB081843. Also described is a polynucleotide (II) encoding  
CC a fusion protein, comprising a sequence selected from 1097 sequences  
CC given in AB065258 to ABP6534 ligated in frame to a polynucleotide  
CC encoding a heterologous polypeptide. (I) has anti-diarrhetic and  
CC antibacterial activities, and can be used as an inhibitor of Salmonella.  
CC (I) (which is a probe) is useful for the detection and/or identification  
CC of Bifidobacterium longum in a biological sample. A carrier containing  
CC the lactic acid bacterium Bifidobacterium longum NCC2705 (NCIM I-2618)  
CC can be used for preventing and/or treating diarrhoea brought about by  
CC pathogenic bacteria and/or rotavirus. The carrier is a food composition  
CC selected from milk, yogurt, curd, cheese, fermented milks, milk based  
CC fermented products, ice-creams, fermented cereal based products, milk  
CC based powders, infant formula, pet food or a pharmaceutical composition  
CC selected from tablets, liquid bacterial suspensions, dried oral  
CC supplement, wet oral supplement, dry tube feeding or wet tube feeding.  
CC (I) is useful in DNA arrays or chips to carry out analysis of the  
CC expression of the Bifidobacterium gene. AB081844 to AB081850 represent  
CC Bifidobacterium related nucleotide sequences given in the sequence  
CC listing from the present invention but not mentioned further within the  
CC specification.  
CC N.B. The sequence data for this patent is not represented in the printed

CC specification but is based on sequence information supplied by the  
CC European Patent Office.  
XX

SQ Sequence 995 AA;

Query Match 1.6%; Score 8; DB 23; Length 995;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 489 SADDVAKA 496  
|||  
Db 40 SADDVAKA 47

RESULT 14  
AAB12246  
ID AAB12246 standard; peptide; 33 AA.

AC AAB12246;

DT 10-NOV-2000 (first entry)

DE Partial sequence of HIV-1 strain DUR gp41 descending helix.

XX HIV-1; AIDS; human immunodeficiency virus type 1; group O HIV;  
KW acquired immunodeficiency syndrome; group M HIV; gp41; DUR.  
XX Human immunodeficiency virus type 1.

OS Human immunodeficiency virus type 1.

PN EPI013766-A2.

PD 28-JUN-2000.

PF 29-NOV-1999; 99EP-0309491.

PR 30-NOV-1998; 98US-0110292.

PR 08-FEB-1999; 99US-0119138.

PR 04-NOV-1999; 99US-0433428.

XX (ORTH-) ORTHO-CLINICAL DIAGNOSTICS INC.

XX De Lays R, Zheng J;

XX WPI; 2000-402205/35.

XX New antigenic peptides and peptide functional derivatives, useful for  
PT detection of antibodies produced in response to human immunodeficiency  
PT virus group O antibodies -

XX Example 4; Fig 4; 52bp; English.

XX The present sequence is a partial gp41 protein of Human Immunodeficiency  
CC Virus Type 1 (HIV-1) strain DUR. HIV is the principle aetiological  
CC agent for acquired immunodeficiency syndrome (AIDS). gp41 is a HIV  
CC envelope protein, and may be used as an antigen for the detection of  
CC antibodies produced in response to HIV infection. DUR is a member of  
CC HIV group O. The present sequence is the descending helix of gp41. This  
CC sequence was used in a sequence homology alignment, which in turn was  
CC used to derive consensus sequence peptides (AAB12259 and AAB12260). The  
CC peptides of AAB12259 and AAB12260 were used as group O replacement  
CC peptides. The peptides of AAB12259 and AAB12260 were used to construct  
CC mosaic gp41 proteins, in which the group M immunodominant region was  
CC replaced by the peptides of AAB12259 or AAB12260. The mosaic gp41  
CC proteins (AAB12261 and AAB12262) would be useful as antigens, which  
CC would be used for the detection of anti-group O HIV antibodies produced  
CC in response to HIV infection.

SQ Sequence. 33 AA;

Query Match 1.4%; Score 7; DB 21; Length 33;  
Best Local Similarity 100.0%; Pred. No. 69;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 467 DOQIANV 473  
DB 1 DOQIANV 7

## RESULT 15

AAAG93319  
ID AAG93319 standard; Protein; 39 AA.

XX AAG93319;

DT 13-SEP-2001 (first entry)

DE Human protein HP10511.

KW Human; gene therapy; tumour.

XX Homo sapiens.

PN WO200142302-A1.

PD 14-JUN-2001.

PF 06-DEC-2000; 2000WO-JP08631.

PR 06-DEC-1999; 99JP-0346863.

PR 08-FEB-1999; 99JP-0346864.

PR 10-FEB-2000; 2000JP-0034090.

PR 10-FEB-2000; 2000JP-0034091.

PR 14-FEB-2000; 2000JP-0035829.

PR 14-FEB-2000; 2000JP-0035829.

PR 14-MAR-2000; 2000JP-0071161.

PR 30-MAY-2000; 2000JP-0160851.

PA (NISC-) JAPAN SCI & TECHNOLOGY CORP.

PI Kato S, Eguchi C, Saeki M;

DR WPI; 2001-381646/40.

DR N-PSDB; AAH68604.

PS Claim 1; Page 416; 471pp; Japanese.

CC The present sequence is a human protein. The human protein, preferably

CC originated from tumour cell line, is applicable as a drug, a reagent for

CC studying intracellular protein networks and a protein source for

CC screening proteins for binding low molecular weight drugs. The human

CC protein coding sequence is useful for gene diagnosis and gene therapy,

CC expression vectors and transformant cells for detection of ligands and

CC receptors.

SO Sequence 39 AA;

QY Query Match 1.4%; Score 7; DB 22; Length 39;

DB Best Local Similarity 100.0%; Pred. No. 80;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 150 GGGAAS 156  
DB 6 GGGAAS 12

RESULT 16  
ID ABG55536 standard; Peptide; 71 AA.

XX ABG55536;  
AC ABG55536;

XX 25-FEB-2003 (first entry) \*

DT Human liver peptide, SEQ ID No 34184.

DE Human liver peptide, SEQ ID No 34184.

XX Human; liver; cirrhosis; hyperlipoproteinaemia; hyperlipidaemia;

XX hypercholesterolaemia; coronary heart disease.

XX Homo sapiens.

PN WO200157273-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00664.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

PA (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

DR WPI; 2001-488898/53.

PT Human genome-derived single exon nucleic acid probes useful for

PT analysing gene expression in human adult liver -

PS Claim 27; SEQ ID No 34184; 658pp; English.

XX The invention relates to a single exon nucleic acid probe (SENP) (I) for

XX measuring human gene expression in a sample derived from human adult

XX liver, comprising one of 13109 defined nucleotide sequences given in the

XX specification (or complements/ fragments). The probe hybridises at high

XX stringency to a nucleic acid molecule expressed in the human adult

XX liver. (I) may be used for predicting, measuring and displaying gene

XX expression in samples derived from human adult liver. The genes

XX identified may be involved in genetic liver diseases such as cirrhosis,

XX hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which

XX is associated with coronary heart disease. ABG47348-ABG5930 represent

XX human liver single exon encoded peptides of the invention.

XX Note: The sequence information for this patent does not appear in the

XX printed specification but was obtained in electronic format directly

XX from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

SO Sequence 71 AA;

QY Query Match 1.4%; Score 7; DB 22; Length 71;

DB Best Local Similarity 100.0%; Pred. No. 1.3e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 494 AKAKTK 500  
DB 28 AKAKTK 34

RESULT 17  
ID ABB40293 standard; Peptide; 71 AA.

XX ABB40293;  
AC ABB40293;  
XX ABB40293;  
DT 04-FEB-2002 (first entry)

DE Peptide #7799 encoded by human foetal liver single exon probe.

XX Human; foetal liver; gene expression; single exon nucleic acid probe.

XX Human; foetal liver; gene expression; single exon nucleic acid probe.

OS Homo sapiens.  
XX  
PN WO200157277-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 30-JAN-2001; 2001WO-US00669.  
XX  
PR 04-FEB-2000; 2000US-0180312.  
PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
DR WPI; 2001-483447/52.  
XX  
PT Human genome-derived single exon nucleic acid probes useful for  
PT analyzing gene expression in human fetal liver -  
XX  
PS Claim 27; SEQ ID NO 32928; 639pp + sequence listing; English.  
XX  
CC The invention relates to a single exon nucleic acid probe for  
CC measuring human gene expression in a sample derived from human foetal  
CC liver. The single exon nucleic acid probes may be used for predicting,  
CC measuring and displaying gene expression in samples derived from human  
CC fetal liver. The present sequence is a peptide encoded by a single exon  
CC nucleic acid probe of the invention.  
CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 71 AA;  
XX  
Query Match 1.4%; Score 7; DB 22; Length 71;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02; Mismatches 0; Indels 0; Gaps 0;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 494 AKAKTK 500  
DB 28 AKAKTK 34  
XX  
RESULT 18  
AAM61086  
ID AAM61086 standard; Protein; 71 AA.  
XX  
AC AAM61086;  
XX  
DT 05-NOV-2001 (first entry)  
XX  
DE Human brain expressed single exon probe encoded protein SEQ ID NO: 33191.  
XX  
KW Human; brain expressed exon; gene expression analysis; probe;  
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;  
XX  
OS Homo sapiens.  
XX  
PN WO200157275-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 30-JAN-2001; 2001WO-US00667.  
XX  
PR 04-FEB-2000; 2000US-0180312.  
PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
XX

PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
DR WPI; 2001-483446/52.  
XX  
PT Single exon nucleic acid probes for analyzing gene expression in human  
PT brains -  
XX  
PS Example 4; SEQ ID NO: 33191; 650pp + Sequence listing; English.  
XX  
CC The present invention provides a number of single exon nucleic acid  
CC probes which are derived from genomic sequences expressed in the human  
CC brain. They can be used to measure gene expression in brain cell samples,  
CC which may enable the diagnosis and improved treatment of nervous system  
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,  
CC epilepsy and cancers. The present sequence is a protein encoded by one of  
CC the probes of the invention.  
XX  
SQ Sequence 71 AA;  
XX  
Query Match 1.4%; Score 7; DB 22; Length 71;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02; Mismatches 0; Indels 0; Gaps 0;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 494 AKAKTK 500  
DB 28 AKAKTK 34  
XX  
RESULT 19  
AAM73789  
ID AAM73789 standard; Protein; 71 AA.  
XX  
AC AAM73789;  
XX  
DT 06-NOV-2001 (first entry)  
XX  
DE Human bone marrow expressed probe encoded protein SEQ ID NO: 34095.  
XX  
KW Human; bone marrow expressed exon; gene expression analysis; probe;  
KW microarray; cancer; Leukaemia; lymphoma; myeloma.  
XX  
OS Homo sapiens.  
XX  
PN WO200157276-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 30-JAN-2001; 2001WO-US00668.  
XX  
PR 04-FEB-2000; 2000US-0180312.  
PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
DR WPI; 2001-488900/53.  
XX  
PT Human genome-derived single exon nucleic acid probes useful for  
PT analyzing gene expression in human bone marrow -  
XX

PS Example 4; SEQ ID NO: 34095; 658bp + Sequence Listing; English.  
 CC The present invention provides a number of single exon nucleic acid  
 CC probes which are derived from genomic sequences expressed in the human  
 CC bone marrow. They can be used to measure gene expression in bone marrow  
 CC samples, which may enable the improved diagnosis and treatment of cancers  
 CC such as lymphoma, leukemia and myeloma. The present sequence is a  
 CC protein encoded by one of the probes of the invention.  
 SQ Sequence 71 AA;

Query Match 1.4%; Score 7; DB 22; Length 71;  
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 494 AKAKTK 500  
 |||||  
 Db 28 AKAKTK 34

RESULT 20  
 AAM33976  
 ID AAM33976 standard; Protein; 71 AA.

AC AAM33976;  
 DT 17-OCT-2001 (first entry)

DE Peptide #8013 encoded by probe for measuring placental gene expression.  
 XX  
 XX Probe; microarray; human; placenta; antenatal diagnosis;  
 KW genetic disorder.

OS Homo sapiens.  
 XX  
 XX WO200157272-A2.

PD 09-AUG-2001.  
 XX  
 XX 30-JAN-2001; 2001WO-US00663.

PF 04-FEB-2000; 2000US-0180312.  
 XX  
 XX 26-MAY-2000; 2000US-0207456.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-0234687.  
 PR 27-SEP-2000; 2000US-0236359.  
 PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-48897/53.

PT Human genome-derived single exon nucleic acid probes useful for  
 XX analyzing gene expression in human placenta -  
 XX  
 XX Claim 27; SEQ ID No 34245; 654bp; English.

XX The present invention relates to single exon nucleic acid probes (SENP:  
 CC see A113135-A157546). The present sequence is a peptide encoded by one  
 CC such probe. The probes are useful for producing a microarray for  
 CC predicting, measuring and displaying gene expression in samples derived  
 CC from human placenta. The probes are useful for antenatal diagnosis of  
 CC human genetic disorders.  
 XX

SQ Sequence 71 AA;

Query Match 1.4%; Score 7; DB 22; Length 71;  
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 494 AKAKTK 500  
 |||||  
 Db 28 AKAKTK 34

RESULT 21  
 ABG43675  
 ID ABG43675 standard; Peptide; 71 AA.

XX ABG43675;

DT 19-AUG-2002 (first entry)

DE Human peptide encoded by genome-derived single exon probe SEQ ID 33340.

XX Human; single exon probe; asthma; lung cancer; COPD; ILD;  
 KW chronic obstructive pulmonary disease; interstitial lung disease;  
 KW familial idiopathic pulmonary fibrosis; neurofibromatosis;  
 KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;  
 KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemostasis;  
 KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;  
 KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;  
 KW primary ciliary dyskinesia; pulmonary hypertension;  
 KW hyaline membrane disease.

XX Homo sapiens.

PN WO200186003-A2.

PD 15-NOV-2001.

PF 30-JAN-2001; 2001WO-US00665.

XX 04-FEB-2000; 2000US-180312P.  
 XX 26-MAY-2000; 2000US-207456P.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-234687P.  
 PR 27-SEP-2000; 2000US-236359P.  
 PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2002-114183/15.

PT Spatially-addressable set of single exon nucleic acid probes, used to  
 XX measure gene expression in human lung samples -  
 XX  
 XX Claim 27; SEQ ID No 33340; 634bp; English.

XX The invention relates to a spatially-addressable set of single exon  
 CC nucleic acid probes for measuring gene expression in a sample derived  
 CC from human lung comprising single exon nucleic acid probes having one of  
 CC 12614 nucleic acid sequences mentioned in the specification, or their  
 CC complements or the 12387 open reading frames derived from the 12614  
 CC probes. Also included are a microarray comprising the novel set of  
 CC probes; the novel set of probes which hybridise at high stringency to a  
 CC nucleic acid expressed in the human lung; measuring gene expression in a  
 CC sample derived from human lung, comprising (a) contacting the array with  
 CC a collection of detectably labeled nucleic acid derived from human lung  
 CC mRNA, and (b) measuring the labeled detectably bound to each probe of  
 CC the array; identifying exons in a eukaryotic genome, comprising  
 CC (a) algorithmically predicting at least one exon from genomic sequences  
 CC of the eukaryote; and (b) detecting specific hybridisation of detectably  
 CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,  
 CC having a fragment identical to the predicted exon, the probe is included  
 CC in the above mentioned microarray; assigning exons to a single gene,  
 CC comprising (a) identifying exons from genomic sequence by the method  
 CC above and (b) measuring the expression of each of the exons in several  
 CC tissues and/or cell types using hybridisation to a single exon  
 CC microarrays having a probe with the exon, where a common pattern of

expression of the exon in the tissues and/or cell types indicates that the exon should be assigned to a single gene; a peptide comprising one of 12011 sequences, mentioned in the specification, or encoded by the probe/open reading frames (ORF). The probes are used for gene expression analysis, and for identifying exons in a gene, particularly using human lung derived mRNA and for the study of lung diseases such as asthma, lung cancer, chronic obstructive pulmonary disease (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease, Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary haemorrhoidosis, pulmonary histiocytosis, lymphangioleiomyomatosis, CC pulmonary alveolar proteinosis, Karsenger syndrome, fibrocystic pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension and hyaline membrane disease. The present sequence is a peptide/protein encoded by a single exon probe of the invention.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://wipo.int/pub/published_pct_sequences).

SO Sequence 71 AA;

Query Match 1.4%; Score 7; DB 23; Length 71;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 494 AKAKTK 500  
| | | | |  
DB 28 AKAKTK 34

RESULT 22  
ID ABB60712 standard; Protein: 90 AA.  
AC ABB60712;  
XX  
XX 26-MAR-2002 (first entry)  
DT  
XX  
DE Drosophila melanogaster polypeptide SEQ ID NO 8928.  
XX  
XX Drosophila; developmental biology; cell signalling; insecticide;  
KM pharmaceutical.  
XX  
XX Drosophila melanogaster.  
OS  
XX  
XX WO200171042-A2.  
FN  
XX  
XX 27-SEP-2001.  
PD  
XX  
XX 23-MAR-2001; 2001WO-US09231.  
PF  
XX  
XX 23-MAR-2000; 2000US-191637P.  
PR  
XX  
XX 11-JUL-2000; 2000US-0614150.  
PT  
XX  
XX (PEKE ) PE GORP NY.  
PA  
XX  
XX  
PI Venter JC, Adams M, Li FWD, Myers EW;  
XX  
XX WPI; 2001-656860/75.  
DR  
XX  
XX N-PSDB; ABL04815.  
PT  
XX  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX  
XX  
PS Disclosure; SEQ ID NO 8928; 21pp + Sequence Listing; English.  
XX  
XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL1840-ABL16175) and the encoded proteins  
CC (ABBS7737-ABBS7072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://wipo.int/pub/published_pct_sequences).  
XX  
XX

SO Sequence 90 AA;

Query Match 1.4%; Score 7; DB 22; Length 90;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 121 EGDTEV 127  
| | | | |  
DB 75 EGDTEV 81

RESULT 23  
ID AAO10927 standard; Protein: 94 AA.  
AC AAO10927;  
XX  
XX 06-NOV-2001 (first entry)  
DT  
XX  
DE Human polypeptide SEQ ID NO 24819.  
XX  
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
KM vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KM tissue growth factor; immunomodulatory; cancer; leukaemia;  
KM nervous system disorders; arthritis; inflammation.  
XX  
XX  
XX Homo sapiens.  
OS  
XX  
XX WO200164835-A2.  
FN  
XX  
XX 07-SEP-2001.  
PD  
XX  
XX 26-FEB-2001; 2001WO-US04927.  
PF  
XX  
XX 28-FEB-2000; 2000US-0515126.  
PR  
XX  
XX 18-MAY-2000; 2000US-0577409.  
PT  
XX  
XX (HYSE-) HYSEQ INC.  
PA  
XX  
XX  
PI Tang YT, Liu C, Drmanac RT;  
XX  
XX WPI; 2001-514838/56.  
DR  
XX  
XX N-PSDB; AAI90858.  
PT  
XX  
XX Isolated nucleic acids and polypeptides, useful for preventing  
PT diagnosing and treating e.g. leukaemia, inflammation and immune  
PT disorders -  
XX  
XX  
PS Claim 20; SEQ ID NO 24819; 1399pp + Sequence Listing; English.  
XX  
XX The invention relates to human polynucleotides (AAI79941-AAI93841) and  
CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activity/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://wipo.int/pub/published_pct_sequences).  
XX  
XX  
SQ Sequence 94 AA;



Query Match 1.4%; Score 7; DB 22; Length 94;  
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 147 TGGGGA 153  
 DB 26 TGGGGA 32

## RESULT 24

AA07245  
 ID AA07245 standard; Protein; 104 AA.

AC AA07245;

DT 09-APR-1997 (first entry)

DE HIV-1 group O strain DUR gp41 immunodominant region peptide.

KW Human immunodeficiency virus; subgroup; strain; AIDS; homology; envelop;

KW gp120; gp41; seropositive; antibody; primer; probe; group O; group M.

OS Human immunodeficiency virus type 1.

PN WO9612809-A2.

PD 02-MAY-1996.

PF 20-OCT-1995; 95WO-FR01391.

PR 03-MAR-1995; 95FR-0002526.

PR 20-OCT-1994; 94FR-0012554.

PA (INSP) INST PASTEUR.

PI Borman A, Charneau P, Clavel F, Cohen JHM, Guetard D;

PI Donjon de Saint-Martin J, Montagnier L, Quillent C;

DR WPI; 1996-230610/23.

DR N-PSDB; AAT42241.

PT New antigenic HIV-1 group O strain proteins and related nucleic

PT acids - useful in diagnosis, vaccines, therapy etc., of infection by

PT HIV-1 group O strains VAV or DUR

PS Claim 24; Fig 12; 108pp; French.

CC The invention relates to the isolation of a novel subgroup of the human

CC immunodeficiency virus (HIV) type 1, designated group O. In particular,

CC the inventors have isolated 2 new strains of the group O virus: strains

CC VAV and DUR. Strain VAV was isolated from a French AIDS patient and has

CC homology to the recently characterised Cameroonian HIV strains AN770 and

CC WPI5180. The DUR strain was isolated from a seropositive patient from

CC the Cameroons who showed atypical seroreactivity. The sequence presented

CC here represents the amino acid sequence of the DUR strain gp41

CC immunodominant region. The DNA and protein sequences are used to

CC generate peptides for detection of antibodies from patients infected

CC with the new group O strains, as well as primers and probes to detect

CC the viral nucleic acids. The peptides and nucleic acid sequences derived

CC from these strains are able to distinguish between the group O and group

CC M viral strains.

Sequence 104 AA;

Query Match 1.4%; Score 7; DB 17; Length 104;

Best Local Similarity 100.0%; Pred. No. 1.9e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 467 DOOIANV 473  
 DB 83 DOOIANV 89

## RESULT 25

ABG26287  
 ID ABG26287 standard; Protein; 105 AA.

AC ABG26287;

DT 18-FEB-2002 (first entry)

DE Novel human diagnostic protein #26278.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;

KW food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

PA (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

PI WPI; 2001-639362/73.

DR N-PSDB; AAS90474.

PT New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostics, forensics, gene mapping, identification of mutations

PT responsible for genetic disorders or other traits and to assess

PT biodiversity

PS Claim 20; SEQ ID No 5646; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and

XX polypeptide (II) sequences. (I) is useful as hybridisation probes,

XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome

XX and gene mapping, and in recombinant production of (II). The

XX polynucleotides are also used in diagnostics as expressed sequence tags

XX for identifying expressed genes. (I) is useful in gene therapy techniques

XX to restore normal activity of (II) or to treat disease states involving

XX (II). (II) is useful for generating antibodies against it, detecting or

XX quantitating a polypeptide in tissue, as molecular weight markers and as

XX a food supplement. (II) and its binding partners are useful in medical

XX imaging of sites expressing (II). (I) and (II) are useful for treating

XX disorders involving aberrant protein expression or biological activity.

XX The polypeptide and polynucleotide sequences have applications in

XX diagnostics, forensics, gene mapping, identification of mutations

XX and to produce other types of data and products dependent on DNA and

XX amino acid sequences. ABG0010-ABG30377 represent novel human

XX Note: The sequence data for this patent did not appear in the printed

XX specification, but was obtained in electronic format directly from WIPO

XX at ftp.wipo.int/pub/published\_pat\_sequences.

Sequence 105 AA;

Query Match 1.4%; Score 7; DB 22; Length 105;

Best Local Similarity 100.0%; Pred. No. 1.9e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 434 HARAASI 440  
 DB 82 HARAASI 88

## RESULT 26

```

AAR72394
ID AAR72394 standard; Protein; 111 AA.
XX
XX AAR72394;
AC
XX
XX 14-DEC-1995 (first entry)
DT
XX
XX Agmenellum quadruplicatum PR-6 rubisco small subunit.
DE
XX
XX Agmenellum quadruplicatum PR-6 rubisco small subunit.
DE
XX
XX Rubisco; ribulose 1,5-bisphosphate carboxylase/oxygenase;
KM large subunit; carbon dioxide fixation; Synechococcus sp.
XX
XX Agmenellum quadruplicatum PR-6.
OS
XX
XX JP07079782-A.
PN
XX
XX 28-MAR-1995.
PD
XX
XX 18-JUN-1993; 93JP-0184304.
PF
XX
XX 18-JUN-1993; 93JP-0184304.
PR
XX
XX (KANT ) KANSAI DENRYOKU KK.
PA
XX
XX WPI; 1995-157852/21.
DR
XX
XX N-PSDB; AAQ87824.
DR
XX
XX Promoter sequence of Synechococcus PCC7002-originated rubisco gene -
PT used in a vector to produce Cyano:bacterium with improved carbon
PT d:oxide fixation
PT
XX
XX Claim 12; Page 25; 28pp; Japanese.
PS
XX
XX A sequence including the ORFs coding for the large and small
CC subunits of the rubisco enzyme has been isolated from Agmenellum
CC quadruplicatum PR-6 (Synechococcus sp. strain PCC7002). The
CC region coding for the rubisco small subunit is claimed; the enzyme
CC is involved in carbon dioxide fixation.
CC
XX
XX Sequence 111 AA;
SQ
Query Match 1.4%; Score 7; DB 16; Length 111;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 465 LSDQQA 471
DB 19 LSDQQA 25

```

```

XX
XX 19-MAY-2000; 2000US-205515P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX
XX Birse CE, Rosen CA;
PI
XX
XX WPI; 2002-122018/16.
DR
XX
XX N-PSDB; ABL89905.
DR
XX
XX Novel 1405 isolated polypeptides, useful for diagnosis, treatment and
PT prevention of neural, immune system, muscular, reproductive,
PT gastrointestinal, pulmonary, cardiovascular, renal and proliferative
PT disorders -
PT
XX
XX Claim 11; SEQ ID NO 1872; 2081pp + Sequence Listing; English.
PS
XX
XX The invention relates to novel genes (ABL89449-ABL90853) and proteins
CC (ABL89040-ABL90444) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC hemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemia;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
CC
XX
XX Sequence 114 AA;
SQ
Query Match 1.4%; Score 7; DB 23; Length 114;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 17 ALMLSGC 23
DB 57 ALMLSGC 63

```

```

RESULT 27
ID ABB89496
XX
XX ABB89496 standard; Protein; 114 AA.
XX
XX ABB89496;
AC
XX
XX 24-MAY-2002 (first entry)
DT
XX
XX Human polypeptide SEQ ID NO 1872.
DE
XX
XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
XX antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;
XX vulnery; anticonvulsant; antibacterial; antifungal; antiparasitic;
XX cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
XX neurological disease; infection; human; secreted protein.
XX
XX Homo sapiens.
OS
XX
XX WO200190304-A2.
PN
XX
XX 29-NOV-2001.
PD
XX
XX 18-MAY-2001; 2001WO-US16450.
PF

```

```

RESULT 28
ID AAB27869
XX
XX AAB27869 standard; Protein; 118 AA.
XX
XX AAB27869;
AC
XX
XX 29-JAN-2001 (first entry)
DT
XX
XX Sequence homologous to protein fragment encoded by gene 27.
DE
XX
XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
XX antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;
XX vulnery; anticonvulsant; antibacterial; antifungal; antiparasitic;
XX cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
XX neurological disease; infection; human; secreted protein.
XX
XX Homo sapiens.
OS
XX
XX WO200055199-A1.
PN
XX
XX 21-SEP-2000.
PD
XX
XX 09-MAR-2000; 2000WO-US06014.
PF
XX
XX 12-MAR-1999; 99US-0124095.
PR
XX
XX 11-JUN-1999; 99US-0138598.

```

PR 03-DEC-1999; 99US-0168665.  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX PI Rosen CA, Ruben SM, Komatsoulis G;  
XX  
XX WPI; 2000-572359/53.  
DR  
XX  
XX PT Isolated nucleic acid molecule encoding a human secreted protein is  
XX used in preventing, treating or ameliorating a medical condition -  
XX  
XX PS Disclosure; Page 417; 433pp; English.  
XX  
XX CC The invention relates to the isolation of genes AAC59215-C59261 encoding  
XX the human secreted proteins AAB27794-B27840. This sequence represents a  
XX peptide fragment homologous to the protein encoded by the gene given  
XX in the descriptor line. The sequence is a search result from a BLASTX  
XX homology search. The genes and proteins are useful for preventing,  
XX ameliorating or treating medical conditions, e.g. by protein or gene  
XX therapy. The genes are isolated from a range of human tissues disclosed  
XX in the specification. The nucleic acids, proteins, antibodies and  
XX (ant)agonists are useful in the diagnosis, treatment and prevention of:  
XX (a) cancer, e.g. breast and ovarian cancer, and other cancers of the  
XX adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver,  
XX lung, or urogenital; (b) immune disorders e.g. Addison's disease,  
XX allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,  
XX diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid  
XX arthritis and ulcerative colitis; (c) cardiovascular disorders such  
XX as myocardial ischaemias; (d) wound healing; (e) neurological diseases  
XX e.g. cerebral anoxia and epilepsy; and (f) infectious diseases such  
XX as viral, bacterial, fungal and parasitic infections.  
XX  
SQ Sequence 118 AA;  
XX  
XX Query Match 1.4%; Score 7; DB 21; Length 118;  
XX Best Local Similarity 100.0%; Pred.No.2.1e+02;  
XX Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
XX  
QY 494 AKAKTK 500  
XX |||||  
XX 87 AKAKTK 93  
XX  
Db  
XX  
XX RESULT 29  
XX ABOG8519  
XX ID ABOG8519 standard; Protein; 124 AA.  
XX  
XX AC ABOG8519;  
XX  
XX DT 13-FEB-2002 (first entry)  
XX  
XX DE Novel human diagnostic protein #8510.  
XX  
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
XX food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
XX OS Homo sapiens.  
XX  
XX PN WO200175067-A2.  
XX  
XX PD 11-OCT-2001.  
XX  
XX PF 30-MAR-2001; 2001WO-US06631.  
XX  
XX PR 31-MAR-2000; 2000US-0540217.  
XX  
XX PR 23-AUG-2000; 2000US-0649167.  
XX  
XX PA (HYSE-) HYSEQ INC.  
XX  
XX PI Drmanac RT, Liu C, Tang YT;  
XX  
XX WPI; 2001-639362/73.  
XX  
XX N-PSDB; AAS72706.  
XX  
DR

XX  
XX PT New isolated polynucleotide and encoded polypeptides, useful in  
XX diagnostics, forensics, gene mapping, identification of mutations  
XX responsible for genetic disorders or other traits and to assess  
XX biodiversity -  
XX  
XX PS Claim 20; SEQ ID No 38878; 103pp; English.  
XX  
XX CC The invention relates to isolated polynucleotide (I) and  
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,  
XX CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
XX and gene mapping, and in recombinant production of (II). The  
XX CC polynucleotides are also used in diagnostics as expressed sequence tags  
XX for identifying expressed genes. (I) is useful in gene therapy techniques  
XX CC to restore normal activity of (II) or to treat disease states involving  
XX CC (II). (II) is useful for generating antibodies against it, detecting or  
XX quantitating a polypeptide in tissue, as molecular weight markers and as  
XX a food supplement. (II) and its binding partners are useful in medical  
XX CC imaging of sites expressing (II). (I) and (II) are useful for treating  
XX CC disorders involving aberrant protein expression or biological activity.  
XX CC The polypeptide and polynucleotide sequences have applications in  
XX CC diagnostics, forensics, gene mapping, identification of mutations  
XX responsible for genetic disorders or other traits to assess biodiversity  
XX and to produce other types of data and products dependent on DNA and  
XX CC amino acid sequences. ABO0010-ABG30377 represent novel human  
XX CC diagnostic amino acid sequences of the invention.  
XX CC Note: The sequence data for this patent did not appear in the printed  
XX CC specification, but was obtained in electronic format directly from WIPO  
XX at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 124 AA;  
XX  
XX Query Match 1.4%; Score 7; DB 22; Length 124;  
XX Best Local Similarity 100.0%; Pred.No.2.2e+02;  
XX Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
XX  
QY 326 PGDYLVLV 332  
XX |||||  
XX 87 PGDYLVLV 93  
XX  
Db  
XX  
XX RESULT 30  
XX AAO01520  
XX ID AAO01520 standard; Protein; 124 AA.  
XX  
XX AC AAO01520;  
XX  
XX DT 06-NOV-2001 (first entry)  
XX  
XX DE Human polypeptide SEQ ID NO 15412.  
XX  
XX KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
XX vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
XX tissue growth factor; immunomodulatory; cancer; leukaemia;  
XX nervous system disorders; arthritis; inflammation.  
XX  
XX OS Homo sapiens.  
XX  
XX PN WO200164835-A2.  
XX  
XX PD 07-SEP-2001.  
XX  
XX PF 26-FEB-2001; 2001WO-US04927.  
XX  
XX PR 28-FEB-2000; 2000US-0515126.  
XX  
XX PR 18-MAY-2000; 2000US-0577409.  
XX  
XX PA (HYSE-) HYSEQ INC.  
XX  
XX PI Tang YT, Liu C, Drmanac RT;  
XX  
XX WPI; 2001-514838/56.  
XX  
XX N-PSDB; AAI81451.  
XX  
DR

XX Isolated nucleic acids and polypeptides, useful for preventing  
PT diagnosing and treating e.g. leukaemia, inflammation and immune  
PT disorders -  
XX  
PS Claim 20; SEQ ID NO 15412; 1399PP + Sequence Listing; English.  
XX  
CC The invention relates to human polynucleotides (AA179941-AA193841) and  
CC the encoded proteins (AA000010-AA01910) that exhibit activity elating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 124 AA;  
XX  
Query Match 1.4%; Score 7; DB 22; Length 124;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 69 APEVPP 75  
DB 70 APEVPP 76  
XX  
RESULT 31  
ID ABB55608 standard; Protein; 130 AA.  
XX  
AC ABB55608;  
XX  
DT 16-MAY-2002 (first entry)  
XX  
DE Lactococcus lactis protein rpsI.  
XX  
DE Lactococcus lactis protein rpsI.  
XX  
KM Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese.  
XX  
OS Lactococcus lactis IL1403.  
XX  
FN FR2807446-A1.  
XX  
PD 12-OCT-2001.  
XX  
PF 11-APR-2000; 2000FR-0004630.  
XX  
PR 11-APR-2000; 2000FR-0004630.  
XX  
PA (INRG ) INRA INST NAT RECH AGRONOME.  
XX  
PI Bolocrine A, Sorokine A, Renault P, Ehrlich SD;  
XX  
DR WPI; 2002-043418/06.  
XX  
PT New nucleotide sequence useful in the identification or Lactococcus  
PT lactis and related species -  
XX  
PS Claim 6; SEQ ID NO 2310; 2504PP; French.  
XX  
CC The present invention is related to a Lactococcus lactis nucleotide  
CC sequence (ABA90521) and related proteins (ABB53300-ABB55621). The  
CC nucleic acid sequence is useful in the detection and/or amplification of  
CC nucleic acid sequence, particularly to identify Lactococcus lactis or  
CC related species. The proteins of the invention are useful for the  
CC biosynthesis or biodegradation of a composition of interest. The  
CC invention helps research in lactic bacteria, particularly useful in the

CC production of yogurt and cheese.  
CC Note: The sequence data for this patent is based on equivalent patent  
CC WO200177334 (published 18-OCT-2001) which is available in electronic  
CC format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 130 AA;  
XX  
Query Match 1.4%; Score 7; DB 23; Length 130;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 447 GKTTVG 453  
DB 26 GKTTVG 32  
XX  
RESULT 32  
ID AAG07374 standard; Protein; 144 AA.  
XX  
AC AAG07374;  
XX  
DT 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 4503.  
XX  
KM Protein identification; signal transduction pathway; metabolic pathway;  
KM hybridisation assay; genetic mapping; gene expression control; promoter;  
KM termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135533.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.

PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137528.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 26-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 02-JUL-1999; 99US-0142154.  
PR 06-JUL-1999; 99US-0142055.  
PR 08-JUL-1999; 99US-0142290.  
PR 09-JUL-1999; 99US-0142803.  
PR 12-JUL-1999; 99US-0142920.  
PR 13-JUL-1999; 99US-0143547.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145813.  
PR 27-JUL-1999; 99US-0145818.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147203.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148655.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 1.4%; Score 7; DB 21; Length 144;  
Best Local Similarity 100.0%; Pred. No. 2.5e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 477 TLN5FCN 483  
Db 81 TLN5FCN 87

RESULT 33  
AAC07373  
ID AAC07373 standard; Protein; 149 AA.  
XX  
AC AAC07373;  
XX  
DT 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 4502.  
XX  
XX Protein identification; signal transduction pathway; metabolic pathway;  
KM hybridisation assay; genetic mapping; gene expression control; promoter;  
KM termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125768.  
PR 25-MAR-1999; 99US-0126264.  
PR 25-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.

PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 02-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142820.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144684.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.

```

PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149920.
PR 23-AUG-1999; 99US-0149920.
PR 25-AUG-1999; 99US-0149920.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157665.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

Query Match 1.4%; Score 7; DB 21; Length 149;  
 Best Local Similarity 100.0%; Pred. No. 2.5e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 477 TLN5FGN 483  
 ID TLN5FGN 92

RESULT 34  
 AAB25190  
 ID AAB25190 standard; Protein; 156 AA.  
 XX  
 AC AAB25190;  
 XX  
 DT 27-NOV-2000 (first entry)  
 XX  
 DE Eucalyptus grandis cell signalling involved protein SEQ ID NO:509.  
 XX  
 KW Eucalyptus grandis; Pinus radiata; Monterey pine; plant; modification;  
 KW plant cell signalling; modulation; transgenic plant; pathogen; growth;  
 KW environmental change; development; cell proliferation; differentiation;  
 KW elongation; survival; disease resistance; nutrient metabolism.  
 XX  
 OS Eucalyptus grandis.  
 XX  
 PN WO200042171-A1.  
 XX  
 PD 20-JUL-2000.  
 XX  
 PF 11-JAN-2000; 2000WO-US00724.  
 XX  
 PR 12-JAN-1999; 99US-0228986.  
 PR 01-NOV-1999; 99US-0162866.  
 XX  
 PA (GENE-) GENESIS RES & DEV CORP LTD.  
 XX  
 XX Strabala TJ, Nieuwenhuizen NJ;  
 XX  
 DR WPI; 2000-476052/41.  
 XX  
 PT Isolated polynucleotide encoding a polypeptide involved in cell  
 PT signaling used for generating transgenic plants with modified responses  
 PT to external signals -  
 XX  
 PS Claim 3; Page 225; 527pp; English.  
 XX

CC AAA79263 to AAA79736 and AAB25100 to AAB25570 represent polynucleotide  
 CC and protein sequences isolated from eucalyptus (Eucalyptus grandis) or  
 CC pine (Pinus radiata also known as Monterey pine). The protein sequences  
 CC are involved in cell signalling. The polynucleotide and protein  
 CC sequences can be used to modify the response of plant cells to external  
 CC signals e.g. environmental changes or pathogens during the growth and  
 CC development of a plant. They can be used to modify cell proliferation,  
 CC differentiation, elongation and survival, resistance to disease and  
 CC nutrient metabolism. Examples of modifications which can be produced are  
 CC altered fruit ripening and senescence of leaves and flowers e.g. to  
 CC delay senescence and prolong the life of cut flowers or enhance  
 CC senescence of reproductive organs to engineer sterile plants. Other  
 CC modifications can be used to delay senescence in selected cell types or  
 CC organs providing fruit and vegetables which have a longer shelf life  
 CC between harvest and consumption, or to decrease branching frequency in  
 CC forest tree species giving long stretches of valuable knot-free clear  
 CC wood which can be used in solid timber furniture and veneers.  
 XX  
 SQ Sequence 156 AA;

Query Match 1.4%; Score 7; DB 21; Length 156;  
 Best Local Similarity 100.0%; Pred. No. 2.6e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 443 NGISGKI 449  
 ID NGISGKI 121

RESULT 35  
 AAB58068  
 ID AAB58068 standard; Protein; 167 AA.  
 XX

AC ABB58068;  
 XX  
 XX 26-MAR-2002 (first entry)  
 XX  
 DE Drosophila melanogaster polypeptide SEQ ID NO 996.  
 XX  
 XX Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical.  
 KW  
 OS Drosophila melanogaster.  
 XX  
 XX WO200171042-A2.  
 XX  
 XX 27-SEP-2001.  
 XX  
 XX 23-MAR-2001; 2001WO-US09231.  
 XX  
 XX 23-MAR-2000; 2000US-191637P.  
 PR 11-JUL-2000; 2000US-0614150.  
 XX  
 XX (PEKE ) PE CORP NY.  
 XX  
 PI Venter JC, Adams M, Li PWD, Myers EW;  
 DR WPI; 2001-656860/75.  
 DR N-PSDB; ABL02171.  
 XX  
 XX New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 XX  
 XX Disclosure; SEQ ID NO 996; 21bp + Sequence Listing; English.  
 PS  
 XX The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA  
 CC sequences (AB101840-AB16175) and the encoded proteins  
 CC (ABBS7737-ABBS72072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 XX  
 XX Sequence 167 AA;  
 SQ  
 Query Match 1.4%; Score 7; DB 22; Length 167;  
 Best Local Similarity 100.0%; Pred. No. 2.8e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 72 VPPPVDR 78  
 DB 118 VPPPVDR 124  
 RESULT 36  
 ABP38126  
 ID ABP38126 standard; Protein; 172 AA.  
 AC  
 AC ABP38126;  
 XX  
 XX 24-JUL-2002 (first entry)  
 XX  
 DE Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:2971.  
 XX  
 XX Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;  
 KW antibacterial; gene therapy.  
 XX  
 OS Staphylococcus epidermidis.  
 XX  
 XX US6380370-B1.  
 XX

PD 30-APR-2002.  
 XX  
 XX 13-AUG-1998; 98US-0134001.  
 PF  
 XX  
 PR 14-AUG-1997; 97US-055779P.  
 PR 08-NOV-1997; 97US-064964P.  
 XX  
 XX (GENO-) GENOME THERAPEUTICS CORP.  
 XX  
 PI Doucette-Stamm LA, Bush D;  
 XX  
 XX WPI; 2002-381255/41.  
 DR N-PSDB; ABN90671.  
 XX  
 XX Novel isolated nucleic acid encoding a Staphylococcus epidermidis  
 PT polypeptide, useful for diagnosing and treating bacterial infections -  
 XX  
 XX Disclosure; SEQ ID 2971; 267bp; English.  
 PS  
 XX ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading  
 CC frame (ORF) nucleic acid sequences which encode the amino acid sequences  
 CC given in ABP35124 to ABP37960. The S. epidermidis sequences have  
 CC antibacterial activity and can be used in gene therapy. The sequences  
 CC can also be used in the diagnosis and treatment of bacterial infections,  
 CC particularly S. epidermidis infections. The sequences can be used to  
 CC screen for compounds able to interfere with the S. epidermidis life  
 CC cycle or inhibit S. epidermidis infection.  
 CC N.B. The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from the  
 CC USPTO web site.  
 CC  
 XX  
 XX Sequence 172 AA;  
 SQ  
 Query Match 1.4%; Score 7; DB 23; Length 172;  
 Best Local Similarity 100.0%; Pred. No. 2.9e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 219 FYTKGRY 225  
 DB 151 FYTKGRY 157  
 RESULT 37  
 AAY28428  
 ID AAY28428 standard; Protein; 180 AA.  
 AC  
 AC AAY28428;  
 XX  
 XX 15-FEB-2000 (first entry)  
 XX  
 DE Wheat branched chain amino acid transaminase amino acid sequence.  
 XX  
 XX Corn; soybean; wheat; rice; dihydroxyacid dehydratase; leuc; leuc;  
 KW branched chain amino acid transferase; biosynthetic enzyme; antibody;  
 KW 3-isopropylmalate dehydratase.  
 XX  
 XX Trilicium aestivum.  
 OS  
 XX  
 XX Key Location/Qualifiers  
 FH Misc-difference 176 /note= "encoded by TAG"  
 FT Misc-difference 179 /note= "encoded by ATA"  
 FT  
 XX  
 XX WO9921880-A2.  
 XX  
 XX 06-MAY-1999.  
 PD  
 XX 20-OCT-1998; 98WO-US22081.  
 PF  
 XX 28-OCT-1997; 97US-0063423.  
 PR  
 XX (DUPO ) DU PONT DE NEMOURS & CO E I.  
 XX



XX Falco SC, Cahoon RE, Hiltz WD, Kinney AJ, Rafalski JA;  
 XX WPI; 2000-022904/02.  
 DR N-PSDB; AAX89454.  
 XX Nucleic acid fragments encoding branched chain amino acid biosynthetic  
 PT enzymes  
 PS Claim 6; Fig 4; 102pp; English.  
 XX  
 CC AAY8418-Y8431 are fragments of corn, soybean, wheat and rice branched  
 CC chain amino acid transaminase, amino acid sequences. Sequences  
 CC AAX89442-X89465 are nucleic acid fragments that encode all or a  
 CC substantial portion of dihydroxyacid dehydratase, a branched chain amino  
 CC acid transferase, a leuc or a leu subunit of 3-isopropylmalate  
 CC dehydratase from wheat, corn, soybean or rice. These enzymes are involved  
 CC in biosynthesis and utilization of branched-chain amino acids. The  
 CC nucleic acid sequences can be used to alter the level of expression of a  
 CC branched chain amino acid biosynthetic enzyme in a host cell. They can  
 CC also be used to obtain a nucleic acid fragment encoding a branched chain  
 CC amino acid biosynthetic enzyme, and to identify inhibitors of a branched  
 CC chain amino acid biosynthetic enzyme. The encoded proteins may be used  
 CC to prepare antibodies for detecting the proteins in situ in cells, or in  
 CC vitro in cell extracts.  
 CC  
 SQ Sequence 180 AA;  
 XX

Query Match 1.4%; Score 7; DB 21; Length 180;  
 Best Local Similarity 100.0%; Pred. No. 3e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 349 KYEGEEN 355  
 |||||  
 DB 58 KYEGEEN 64

RESULT 38  
 AAM86749  
 ID AAM86749 standard; Protein; 180 AA.  
 XX  
 AC AAM86749;  
 XX

DT 07-NOV-2001 (first entry)

DE Human immune/haematopoietic antigen SEQ ID NO:14342.

XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;  
 KW cytostatic; gene therapy; vaccine; metastasis.  
 XX

OS Homo sapiens.

XX WO200157182-A2.  
 PN

XX 09-AUG-2001.  
 PD

PF 17-JAN-2001; 2001WO-US01354.  
 XX

PR 31-JAN-2000; 2000US-0179065.  
 PR 04-FEB-2000; 2000US-0180628.  
 PR 24-FEB-2000; 2000US-0184664.  
 PR 02-MAR-2000; 2000US-0186350.  
 PR 16-MAR-2000; 2000US-0189874.  
 PR 17-MAR-2000; 2000US-0190076.  
 PR 18-APR-2000; 2000US-0198123.  
 PR 19-MAY-2000; 2000US-0205515.  
 PR 07-JUN-2000; 2000US-0209467.  
 PR 28-JUN-2000; 2000US-0214886.  
 PR 30-JUN-2000; 2000US-0215135.  
 PR 07-JUL-2000; 2000US-0216647.  
 PR 07-JUL-2000; 2000US-0216880.  
 PR 11-JUL-2000; 2000US-0217487.  
 PR 11-JUL-2000; 2000US-0217496.

PR 14-JUL-2000; 2000US-0218290.  
 PR 26-JUL-2000; 2000US-0220963.  
 PR 14-AUG-2000; 2000US-0220964.  
 PR 14-AUG-2000; 2000US-0224518.  
 PR 14-AUG-2000; 2000US-0224519.  
 PR 14-AUG-2000; 2000US-0225213.  
 PR 14-AUG-2000; 2000US-0225214.  
 PR 14-AUG-2000; 2000US-0225266.  
 PR 14-AUG-2000; 2000US-0225267.  
 PR 14-AUG-2000; 2000US-0225268.  
 PR 14-AUG-2000; 2000US-0225270.  
 PR 14-AUG-2000; 2000US-0225447.  
 PR 14-AUG-2000; 2000US-0225757.  
 PR 14-AUG-2000; 2000US-0225758.  
 PR 14-AUG-2000; 2000US-0225759.  
 PR 18-AUG-2000; 2000US-0226279.  
 PR 22-AUG-2000; 2000US-0226681.  
 PR 22-AUG-2000; 2000US-0226686.  
 PR 22-AUG-2000; 2000US-0227182.  
 PR 23-AUG-2000; 2000US-0227182.  
 PR 30-AUG-2000; 2000US-0228924.  
 PR 01-SEP-2000; 2000US-0229287.  
 PR 01-SEP-2000; 2000US-0229343.  
 PR 01-SEP-2000; 2000US-0229344.  
 PR 01-SEP-2000; 2000US-0229345.  
 PR 05-SEP-2000; 2000US-0229509.  
 PR 05-SEP-2000; 2000US-0229513.  
 PR 06-SEP-2000; 2000US-0230437.  
 PR 08-SEP-2000; 2000US-0230438.  
 PR 08-SEP-2000; 2000US-0231242.  
 PR 08-SEP-2000; 2000US-0231243.  
 PR 08-SEP-2000; 2000US-0231244.  
 PR 08-SEP-2000; 2000US-0231413.  
 PR 08-SEP-2000; 2000US-0231414.  
 PR 08-SEP-2000; 2000US-0232080.  
 PR 08-SEP-2000; 2000US-0232081.  
 PR 12-SEP-2000; 2000US-0231968.  
 PR 14-SEP-2000; 2000US-0232397.  
 PR 14-SEP-2000; 2000US-0232398.  
 PR 14-SEP-2000; 2000US-0232399.  
 PR 14-SEP-2000; 2000US-0232400.  
 PR 14-SEP-2000; 2000US-0232401.  
 PR 14-SEP-2000; 2000US-0233063.  
 PR 14-SEP-2000; 2000US-0233064.  
 PR 14-SEP-2000; 2000US-0233065.  
 PR 21-SEP-2000; 2000US-0234223.  
 PR 21-SEP-2000; 2000US-0234274.  
 PR 25-SEP-2000; 2000US-0234997.  
 PR 25-SEP-2000; 2000US-0234998.  
 PR 26-SEP-2000; 2000US-0235484.  
 PR 27-SEP-2000; 2000US-0235834.  
 PR 27-SEP-2000; 2000US-0235836.  
 PR 29-SEP-2000; 2000US-0236327.  
 PR 29-SEP-2000; 2000US-0236327.  
 PR 29-SEP-2000; 2000US-0236367.  
 PR 29-SEP-2000; 2000US-0236368.  
 PR 29-SEP-2000; 2000US-0236369.  
 PR 29-SEP-2000; 2000US-0236370.  
 PR 02-OCT-2000; 2000US-0236802.  
 PR 02-OCT-2000; 2000US-0237037.  
 PR 02-OCT-2000; 2000US-0237038.  
 PR 02-OCT-2000; 2000US-0237039.  
 PR 13-OCT-2000; 2000US-0237040.  
 PR 13-OCT-2000; 2000US-0239335.  
 PR 13-OCT-2000; 2000US-0239337.  
 PR 20-OCT-2000; 2000US-0240960.  
 PR 20-OCT-2000; 2000US-0241221.  
 PR 20-OCT-2000; 2000US-0241785.  
 PR 20-OCT-2000; 2000US-0241786.  
 PR 20-OCT-2000; 2000US-0241787.  
 PR 20-OCT-2000; 2000US-0241806.  
 PR 20-OCT-2000; 2000US-0241809.  
 PR 20-OCT-2000; 2000US-0241826.  
 PR 01-NOV-2000; 2000US-0244617.

PR 08-NOV-2000; 2000US-0246474.  
 PR 08-NOV-2000; 2000US-0246475.  
 PR 08-NOV-2000; 2000US-0246476.  
 PR 08-NOV-2000; 2000US-0246477.  
 PR 08-NOV-2000; 2000US-0246478.  
 PR 08-NOV-2000; 2000US-0246523.  
 PR 08-NOV-2000; 2000US-0246524.  
 PR 08-NOV-2000; 2000US-0246525.  
 PR 08-NOV-2000; 2000US-0246526.  
 PR 08-NOV-2000; 2000US-0246527.  
 PR 08-NOV-2000; 2000US-0246528.  
 PR 08-NOV-2000; 2000US-0246532.  
 PR 08-NOV-2000; 2000US-0246609.  
 PR 08-NOV-2000; 2000US-0246610.  
 PR 08-NOV-2000; 2000US-0246611.  
 PR 08-NOV-2000; 2000US-0246613.  
 PR 17-NOV-2000; 2000US-0249207.  
 PR 17-NOV-2000; 2000US-0249208.  
 PR 17-NOV-2000; 2000US-0249209.  
 PR 17-NOV-2000; 2000US-0249210.  
 PR 17-NOV-2000; 2000US-0249211.  
 PR 17-NOV-2000; 2000US-0249212.  
 PR 17-NOV-2000; 2000US-0249213.  
 PR 17-NOV-2000; 2000US-0249214.  
 PR 17-NOV-2000; 2000US-0249215.  
 PR 17-NOV-2000; 2000US-0249216.  
 PR 17-NOV-2000; 2000US-0249217.  
 PR 17-NOV-2000; 2000US-0249218.  
 PR 17-NOV-2000; 2000US-0249244.  
 PR 17-NOV-2000; 2000US-0249245.  
 PR 17-NOV-2000; 2000US-0249264.  
 PR 17-NOV-2000; 2000US-0249265.  
 PR 17-NOV-2000; 2000US-0249297.  
 PR 17-NOV-2000; 2000US-0249299.  
 PR 17-NOV-2000; 2000US-0249300.  
 PR 01-DEC-2000; 2000US-0250160.  
 PR 01-DEC-2000; 2000US-0250391.  
 PR 05-DEC-2000; 2000US-0251030.  
 PR 05-DEC-2000; 2000US-0251988.  
 PR 05-DEC-2000; 2000US-0256719.  
 PR 06-DEC-2000; 2000US-0251479.  
 PR 08-DEC-2000; 2000US-0251856.  
 PR 08-DEC-2000; 2000US-0251868.  
 PR 08-DEC-2000; 2000US-0251869.  
 PR 08-DEC-2000; 2000US-0251989.  
 PR 08-DEC-2000; 2000US-0251990.  
 PR 11-DEC-2000; 2000US-0254057.  
 PR 05-JAN-2001; 2001US-0259678.  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Rosen CA, Barash SC, Ruben SM;  
 XX  
 DR WPI; 2001-483426/52.  
 DR N-PSDB; AAK59530.  
 XX  
 PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
 PT useful for preventing, diagnosing and/or treating cancers and  
 PT metacasts -  
 XX  
 XX  
 XX Claim 11; SEQ ID NO 14342; 3071pp + Sequence Listing; English.  
 XX  
 CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)  
 CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytotoxic  
 CC activity, and can be used in gene therapy and vaccine production. (I)  
 CC proteins and polynucleotides may be used in the prevention, diagnosis and  
 CC treatment of diseases associated with inappropriate (I) expression. For  
 CC example, they may be used to treat disorders associated with decreased  
 CC expression by rectifying mutations or deletions in a patient's genome  
 CC that affect the activity of (I) by expressing inactive proteins or to  
 CC supplement the patient's own production of (I). Additionally, (I)  
 CC polynucleotides may be used to produce the secreted (I), by inserting  
 CC the nucleic acids into a host cell and culturing the cell to express the

CC protein. (I) proteins and polynucleotides may be used to prevent,  
 CC diagnose and treat immune/haematopoietic-related diseases, especially  
 CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703  
 CC to AAK87694 represent human immune/haematopoietic antigen genomic  
 CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169  
 CC represent sequences used in the exemplification of the present invention.  
 XX  
 SQ Sequence 180 AA;

Query Match 1.4%; Score 7; DB 22; Length 180;  
 Best Local Similarity 100.0%; Pred. No. 3e+02; 0; Indels 0; Gaps 0;  
 Matches 7; Conservative 0; Mismatches 0;

Qy 277 GPNLTSS 283  
 Db 7 GPNLTSS 13

RESULT 39  
 ABB71959  
 ID ABB71959 standard; Protein; 185 AA.  
 XX

AC ABB71959;

DT 26-MAR-2002 (first entry)

XX Drosophila melanogaster polypeptide SEQ ID NO 42669.

KW Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical.

OS Drosophila melanogaster.

PN W0200171042-A2.

XX 27-SEP-2001.

PD 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX (PEKE ) PE CORP NY.

PI Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.  
 DR N-PSDB; ABL16062.

XX

PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -

PS Disclosure; SEQ ID NO 42669; 21pp + Sequence Listing; English.

CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL161840-ABL16175) and the encoded proteins  
 CC (ABBS7737-ABBS7072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX

SQ Sequence 185 AA;

Query Match 1.4%; Score 7; DB 22; Length 185;  
 Best Local Similarity 100.0%; Pred. No. 3e+02; 0; Indels 0; Gaps 0;  
 Matches 7; Conservative 0; Mismatches 0;

QY 225 YGEOGLQ 231  
 |||||  
 Db 138 YGEOGLQ 144

## RESULT 40

ABB71208  
 ID ABB71208 standard; Protein; 187 AA.

AC ABB71208;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 40416.

KM Drosophila; developmental biology; cell signalling; insecticide;  
 pharmaceutical.

OS Drosophila melanogaster.

PN WO200171042-A2.

PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US09231.

PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

PA (PEKE ) PE CORP NY.

PI Venter JC, Adams M, Li PWD, Myers EW;

PI N-PSDB; ABL15311.

DR WPI; 2001-656860/75.

PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 genes from Drosophila and for elucidating cell signalling and cell-cell  
 interactions -

PS Disclosure; SEQ ID NO 40416; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from Drosophila. The invention is

CC useful in developmental biology and in elucidating cell signalling and

CC cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA

CC sequences (ABL1840-ABL16175) and the encoded proteins

CC (ABB57737-ABB72072).

CC The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published\_pct\_sequences.

SQ Sequence 187 AA;

QY 445 LSGKITV 451  
 |||||  
 Db 56 LSGKITV 62

Search completed: August 27, 2003, 18:42:34  
 Job time : 110 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: September 7, 2003, 00:14:41 / Search time 95 Seconds  
(without alignments)  
2332.361 Million cell updates/sec

Title: US-10-088-045-2  
Perfect score: 2626  
Sequence: 1 MSKPTLIKTKTLCALSLML.....NKGGLSADVDVAKAKKTKPN 502

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODE=frame+ p2n.model -DEV=xlh  
-O/cgnt2\_1/USPTO.spool/US10088045/runtat\_04092003\_083143\_3945/app\_query.fasta.1.647  
-De-Issued\_Patents\_NA -OPMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITs=bits -START=1 -END=-1 -MATRIX=blotsum62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=40  
-MODE=LOCAL -OUTPMT=ptc -NORM=ext -HAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US10088045 @CCN 1.1 56 @runtat\_04092003\_083143\_3945 -NCPU=6 -ICPU3  
-NO MAP -LARGEOUTRY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGCLO  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- 1: /cgnt2\_6/prodata/2/ina/5A.COMB.seq.\*
- 2: /cgnt2\_6/prodata/2/ina/5B.COMB.seq.\*
- 3: /cgnt2\_6/prodata/2/ina/5C.COMB.seq.\*
- 4: /cgnt2\_6/prodata/2/ina/5D.COMB.seq.\*
- 5: /cgnt2\_6/prodata/2/ina/5E.COMB.seq.\*
- 6: /cgnt2\_6/prodata/2/ina/backfile1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	163.5	6.2	1938	4	US-09-252-991A-8682
2	163.5	6.2	1971	4	US-09-252-991A-8564
3	163	6.2	1329	3	US-09-252-991A-8564
4	163	6.2	1437	3	US-09-252-991A-8564
5	163	6.2	4830	3	US-09-252-991A-8564
6	158.5	6.0	4403765	3	US-09-103-840A-2
7	158.5	6.0	4411529	3	US-09-103-840A-2
8	155	5.9	1302	4	US-09-252-991A-6541
9	154.5	5.9	963	4	US-09-252-991A-6541
10	153	5.8	1470	4	US-09-252-991A-9169
11	153	5.8	1764	4	US-09-252-991A-8797
12	153	5.8	2256	4	US-09-252-991A-8910

13	152	5.8	1404	1	US-07-985-458-2	Sequence 2, Appli
14	151.5	5.8	1353	4	US-09-252-991A-765	Sequence 765, App
15	151.5	5.8	1524	4	US-09-252-991A-737	Sequence 737, App
16	151.5	5.8	1641	4	US-09-252-991A-646	Sequence 646, App
17	150	5.7	1554	4	US-09-252-991A-4893	Sequence 4893, App
18	150	5.7	2139	4	US-09-252-991A-4910	Sequence 4910, App
19	147.5	5.6	1977	4	US-09-252-991A-8339	Sequence 8339, App
20	146	5.6	1782	4	US-09-252-991A-14102	Sequence 14102, A
21	141	5.4	1588	2	US-08-706-037-24	Sequence 24, Appli
22	141	5.4	1588	2	US-09-005-137-13	Sequence 13, Appli
23	141	5.4	1672	1	US-08-172-331B-13	Sequence 13, Appli
24	136.5	5.2	1960	4	US-09-615-192A-208	Sequence 208, App
25	132.5	5.0	1203	4	US-09-252-991A-8410	Sequence 8410, App
26	132	5.0	2016	4	US-09-328-352-1877	Sequence 1877, App
27	131.5	5.0	1908	4	US-09-328-352-13903	Sequence 13903, App
28	130.5	5.0	1323	4	US-09-252-991A-14169	Sequence 14169, A
29	130	5.0	3641	1	US-08-030-096-5	Sequence 5, Appli
30	130	5.0	6885	3	US-08-746-111-4	Sequence 4, Appli
31	128.5	4.9	6909	2	US-08-804-196-1	Sequence 1, Appli
32	128.5	4.9	6909	2	US-08-804-196-1	Sequence 1, Appli
33	128.5	4.9	6909	3	US-08-746-111-26	Sequence 26, Appli
34	128.5	4.9	6909	4	US-08-454-353A-1	Sequence 1, Appli
35	128.5	4.9	6909	4	US-09-165-019-1	Sequence 1, Appli
36	125.5	4.8	933	4	US-09-252-991A-6517	Sequence 6517, App
37	120.5	4.6	7032	3	US-09-324-867-1	Sequence 867, App
38	119	4.5	1944	4	US-09-252-991A-8062	Sequence 8062, App
39	118.5	4.5	477	4	US-09-252-991A-6506	Sequence 6506, App
40	118.5	4.5	1494	4	US-09-252-991A-1456	Sequence 1456, App
41	118.5	4.5	1536	4	US-09-252-991A-1567	Sequence 1567, App
42	117.5	4.5	4665	4	US-09-297-937C-12	Sequence 12, Appli
43	117	4.5	3196	4	US-09-704-449-1	Sequence 1, Appli
44	114.5	4.4	1791	4	US-09-468-578-5	Sequence 5, Appli
45	114.5	4.4	1791	4	US-09-218-702-1	Sequence 1, Appli

# ALIGNMENTS

RESULT 1  
US-09-252-991A-8682  
; Sequence 8682, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: MARC J. RUBENFIELD et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252, 991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074, 788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094, 190  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 8682  
; LENGTH: 1938  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-8682

# Alignment Scores:

Pred. No.: 1.99e-07 Length: 1938  
Score: 163.50 Matches: 93  
Percent Similarity: 33.96% Conservative: 52  
Best Local Similarity: 21.78% Mismatches: 141  
Query Match: 6.23% Indels: 141  
DB: 4 Gaps: 21

US-10-088-045-2 (1-502) x US-09-252-991A-8682 (1-1938)

OY 19 MetLeuSerGlyCysSerSerGlnAlaAspIysAla---AlaGlnProIysSerSerThr 37  
DB 121 GTTCTGCTGGGCTGTCACGGCGGCTTCAGGCAATTTGCCAGCAAGGCTCCGGACG 180

```

QY 38 ValAspAlaAlaAlaLeuThrAlaAsnAlaAspAsnAlaAlaSerGlnGlnIleGlnGly 57
DB 181 -----CGCGGT 186
QY 58 GluLeuProValIle-----AspAlaIleValThrIleAlaProGluVal 72
DB 187 GCCGCCCTGTGCGAAGGCGACCGAGTTGCACTTGATCGCCGAATGCCCGGTG--- 243
QY 73 ProProValAspArgAspHisProAlaValValValIleValIleValIleValIle 92
DB 244 -----AATTCAAGGCGAAGCCAGCGTGGCTAG--- 273
QY 93 LysValMetArgLeuAlaAspGlyValGlnTyrglnPheTrpThrPheGlyGlnVal 112
DB 274 -----ACCATCAACGGCTCGCTG 291
QY 113 ProGlyIleMetIleArgValArgGlnGlyAspThrIleGluValGlnPheSerAsnHis 132
DB 292 CCGGCACCGACATTGCGTGGCGAGGCGTACACCGTCAACATCCGTGAACCAAC--- 348
QY 133 ProAspSerLysMetProHisAsnValAspPheHisAlaIleThrGlyPro----- 149
DB 349 ---AAGTGGGTGAAGCCACATTCATTCATGCGACGGGATCATCTCGGTATCAGATG 405
QY 150 GlyGlyGlyAlaGlnAlaSerPheThr-----AlaProGlyHisThrSerThrPheSer 167
DB 406 GATGGCGGTGGCGGCATCAGTTCAATGGCATCGCTCCCGCGAGCATCTTCAGTACGCG 465
QY 168 PheLysValAlaLeuGlnProGlyLeuTyrglnValIleIleGlyValAlaProValGlyMet 187
DB 466 TTCAAGGTTGACGACGAGCGGTCTCTACTGTGACAC-----TCGCACTCGCGGCTTC 516
QY 188 HisIleAlaAsnGlyMetTyrglyLeuIleLeuValGlnProLysGlnGlyLeuProLys 207
DB 517 CAGGAATCTCAGCGGCATGTACGAGGCGCATCATCTCAACCCC---GCTGGCGCGAAGAC 573
QY 208 Val-----AspLysGlnTyrglyValMet----- 215
DB 574 ATTCACGCGCGATCGTGAACACGATGCTGTTCGGAATTGAGACGACGAAGACCGATG 633
QY 216 -----GlnGlyAspPheTythr----- 221
DB 634 CGGGTGTTCACCAAGTCAAGTCCCAAGCGCATTTATCAACATCAACCAACCTACCGTC 693
QY 222 -----LysGly 223
DB 694 TTGCATTTCTCCGGGACGCTCGCGAAGGACGCGCGCTCGATGACACAGCGCAG 753
QY 224 LysTyrglyGlnGlnGlyLeuGlnProPheAspMetGlnLysAlaIleArgGluAspAla 243
DB 754 ATGTGAACGAGATCGGATGATCCGACGATCTCGAATGTGTGCGACACGCTG 813
QY 244 GlnTyrglyValAlaPheAsnGlySerValGlyAlaLeuThrGlyGluAsnAlaLeuValAla 263
DB 814 ACCTATCTGGCC---AAGCGGTCACCCCGCC---GGCAATTGAACGCGCTGTC 864
QY 264 LysValGlyGlnThrValArgLeuPheValGlyValGlnGlyProAsnLeuThrSerSer 283
DB 865 CGACCGGCTAGCGCGCTGCGGCTCGCATGCTCAACGCTCGGCGACACCTTACAGC 924
QY 284 PheHisValIleGlyGlnIlePheAspLysValHisPheGlnGlyGlyGlyGluAsn 303
DB 925 GTACGATTCAGAGACTGAAGTCAAGTCAAGTGTGTCAGTGTG--- 969
QY 304 HisAsnIleGlnThrThrLeuIleProAlaGlyAlaAlaIleThrGlnPheLysVal 323
DB 970 -----GTT 972
QY 324 AspValProGlyAspTyrglyValLeuValAspHisAlaIlePheArgAlaPheAsnLysGly 343
DB 973 GACGTG-----GAGCGGCTGACCGTCAATGAG---TTCCGA---TTGCGTCTCGGC 1017
QY 344 ---AlaLeuGlyIleLeuLysValGlnGlyGluAsnHisGlnIleTyrglySerHisLys 362

```

```

DB 1018 GAAACCGTCGACGTGATCGTCCAGCCAGCGGATGACGCTTACAGATCTTCCGCCAGGCG 1077
QY 363 GlnThrAspAlaValAlaTyrglyLeuProGlyGlnGlyAlaProGlnAlaIleAspThrGlnGluAla 382
DB 1078 ATGACCGGACGACGCTAC-----GCGGTGCCACGCGTGCCTGCGTGAAGGT 1125
QY 383 ProLysThrProAlaProAla 389
DB 1126 CTTCAGCTCCCGTGCCTGCC 1146

RESULT 2
US-09-252-991A-8564/c
; Sequence 8564, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 8564
; LENGTH: 1971
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-8564

```

```

Alignment Scores:
Pred. No.: 2,04e-07 Length: 1971
Score: 163.50 Matches: 93
Percent Similarity: 33.96% Conservative: 52
Best Local Similarity: 21.78% Mismatches: 141
Query Match: 6.23% Indels: 141
DB: 4 Gaps: 21

```

US-10-088-045-2 (1-502) x US-09-252-991A-8564 (1-1971)

```

QY 19 MetLeuSerGlyCysSerAsnGlnAlaAspLysAla---AlaGlnProLysSerSerThr 37
DB 1956 GTTCTGCTGCGGTGCTCCAGCCGCGCTTCAAGGATTTGCCAGACAGCTCGGAGCG 1897
QY 38 ValAspAlaAlaAlaLysThrAlaAsnAlaAspAsnAlaAlaSerGlnGlnGly 57
DB 1896 -----CGCGGT 1891
QY 58 GluLeuProValIle-----AspAlaIleValThrIleAlaProGluVal 72
DB 1890 GCCGCCCTGTGCGAAGGCGACCGAGTTGCACTTGATCGCCGAATGCCCGGTG--- 1834
QY 73 ProProValAspArgAspHisProAlaValValValIleValIleValIleValIle 92
DB 1833 -----AATTCAAGGCGAAGCCAGCGTGGCTAG--- 1804
QY 93 LysValMetArgLeuAlaAspGlyValGlnTyrglnPheTrpThrPheGlyGlnVal 112
DB 1803 -----ACCATCAACGGCTCGCTG 1786
QY 113 ProGlyIleMetIleArgValArgGlnGlyAspThrIleGluValGlnPheSerAsnHis 132
DB 1785 CCGGCACCGACATTGCGTGGCGAGGCGTACACCGTCAACATCCGTGAACCAAC--- 1729
QY 133 ProAspSerLysMetProHisAsnValAspPheHisAlaIleThrGlyPro----- 149
DB 1728 ---AAGTGGGTGAAGCCACATTCATTCATGCGACGGGATCATCTCGGTATCAGATG 1672
QY 150 GlyGlyGlyAlaGlnAlaSerPheThr-----AlaProGlyHisThrSerThrPheSer 167

```

```
Db 1671 GATGGCGTCCGGGACATCAAGCTTCAATGCGATCGCTCCGGGACAGCTTCAAGTACCGG 1612
Qy 168 PheYsAlaLeuGlnProGlyLeuTyYrValTYrHisCysAlaValAlaProValGlyMet 187
1611 TTCAAGGTTACAGAGCGGCTCTACTGTTACAC-----TCGCACTCCGGCTTC 1561
Qy 188 HisIleAlaAsnGlyMetYrGlyLeuIleLeuValGlnProLysGlyLeuProLys 207
1560 CAGGAATCAGCGGCGCATGTACGGGGCGCATCATCATCAGACCC--GCTGGCCGCGAGACC 1504
Qy 208 Val-----AspLysGlyTyYrValMet----- 215
1503 ATCCAGCCGCGGATGACCAAGTGTGCTGTTCTCGGANTGACCGACGACGACCGGATG 1444
Qy 216 -----GlnGlyAspPheTyYrThr----- 221
1443 CGGGTTTACCAAGCTCAAGTCCCAAGCAAGCTATTACACTACACCAAGCTACCGTC 1384
Qy 222 -----LysGly 223
1383 TTGCACTTCTCCGGGACGCTTCGCGGACGACGCGCGCTGCGATCGACCAAGCGCAAG 1324
Qy 224 LysTYrGlyGlnGlnGlyLeuGlnProPheAspMetGlyValAlaIleArgGlnAspAla 243
1323 ATGTGACACGATGCGGATGAATCCGACGATCTGCGAGATCTGTCCGACACGCTG 1264
Qy 244 GlyTYrValValPheAsnGlySerValGlyAlaLeuThrGlyGlnAsnAlaLeuLysAla 263
1263 ACCTATCTGCCC--AAGCGCGTCACCCCCG--GCGAATTGACGCGCGCTTC 1213
Qy 264 LysValGlyGlnThrValArgLeuPheValGlyAsnGlyGlyProAsnLeuThrSerSer 283
1212 CAGCGCGGTGAGCGCGCTGCGCGCTGCGCATGTCACGCGCGGCAACCTTCTACGAC 1153
Qy 284 PheHisValIleGlyGlyLeuPheAspLysValHisPheGlyGlyLysGlyLysGlyLys 303
1152 GTACGCACTTCAGGACTGAACCTCAGCGGTGTCACGTCGATGT----- 1108
Qy 304 HisAsnIleGlnThrThrLeuIleProAlaGlyAlaAlaIleThrGlnPheLysVal 323
1107 -----GTT 1105
Qy 324 AspValProGlyAspTYrValLeuValAspHisAlaIlePheArgAlaPheAsnLysGly 343
1104 GACGTG-----GACGCGGTGACCGTCGATGAG-----TTCCGA--TTCCGTCCTGCG 1060
Qy 344 ---AlaLeuGlyIleLeuLysValGlnGlyGlnAsnHisGlyLeuTyrSerHisLys 362
1059 GAACCGTGCAGCTGATCGTCCAGCCGCGGATGACGCTTACGATCTTCCGCCGAGCG 1000
Qy 363 GlnThrAspAlaValTYrLeuProGlnGlyAlaProGlnAlaIleAspThrGlnGlnAla 382
999 ATGAGCCGACAGGCTAC-----GCGCGTCCACGCTGCGCGCTGAGGT 952
Qy 383 ProLysThrProAlaProAla 389
951 CTTCAAGCTCCCGTCCCTGCC 931
Db
RESULT 3
US-09-296-284-23
; Sequence 23, Application US/09296284A
; Patent No. 6204040
; GENERAL INFORMATION:
; APPLICANT: Choi, Bui-Sung
; APPLICANT: Rhee, Sang-Ki
; APPLICANT: Lee, Eun-Hae
; TITLE OF INVENTION: Glucanobacter Suboxydans Sorbitol Dehydrogenase, Genes
; FILE REFERENCE: 1533.0870000
; CURRENT APPLICATION NUMBER: US/09/296,284A
; NUMBER OF SEQ ID NOS: 87
; SOFTWARE: PatentIn Ver. 2.0
```

```
; SEQ ID NO 23
; LENGTH: 1329
; TYPE: DNA
; ORGANISM: Glucanobacter suboxydans
US-09-296-284-23
Alignment Scores:
Pred. No.: 1,25e-07 Length: 1329
Score: 163.00 Matches: 118
Percent Similarity: 35.85% Conservative: 34
Best Local Similarity: 27.83% Mismatches: 144
Query Match: 6.21% Indels: 120
DB: Gaps: 25
US-10-088-045-2 (1-502) x US-09-296-284-23 (1-1329)
Qy 171 LeuGlnProGlyLeuTyYrValTYrHis-----CysAlaValAlaPro 184
22 ATTCAGCGCGGTGCTTACGTCGCGCGCGCTGTCTGACTGCGTTCGCTGACATACCGACTA 81
Qy 185 ValGlyMetHisIleAlaAsnGlyMet-----TYrGly 195
82 CAGCGGACGCTTTGCTGCTGCTGCTGCTGAGATCAAGAGCCGATCGGACAGATCTCTCC 141
Qy 196 Leu---IleLeuValGlnProLysGlyLeuProLysValAspLysGlyTYrVal 214
142 ACCAATATCAGCGCTGACCGCGAATACGGTATCGGCAC--TATACA 186
Qy 215 MetGlnGlyAspPheTYrThrLysGlyLysTYrGlyGlnGlnGlyLeuGlnProPheAsp 234
187 CTCGAAGAT--TTCCGAAGGCG----- 207
Qy 235 MetGlyValAlaIleArgGlyAsp----- 242
208 ATCCGTAAGGATATCCGCAAGACGCGCGGACGCTTATCCGCGCATGCCATCTCTGAG 267
Qy 243 -----AlaGlyTYrValValPheAsnGlySerVal 252
268 TTCGTCGCGCTGTGATGACGACATCAAGGCCATGTATGCTTTCATGATGCGCTG 327
Qy 253 GlyAla-----LeuThrGlyLysAsnAlaLeu 261
328 AA-GCGCGTCCGCTTCAGAACAGCGCGGACATCTTCGCGCATACAGATGGCGTG 386
Qy 262 LysAlaLysValGlyGlnThrValArg--LeuPheValGlyAsnGlyLysProAsnLeu 281
387 -----GCCGTGGCCATGTGCGCGCGCATGTTGTTCCGACTGTACACCGAGCTTCA 440
Qy 281 rSerSerPheHis-----ValIleGlyLysIlePheAspLysValHisPheG 297
441 CAAGAGCATCTCGATCCGGAAGTGCGCGCTGCGCAATACCTC-----GTGAATGGCC 494
Qy 297 uGlyGlyLysGlyGlnAsnHisAsnIleGlnThrLeuIleProAlaGlyAlaAla 317
495 AGCGCATGTGTGCGAGTGCATACG-----CCCGGTGCAATGGCCAT 536
Qy 317 allerThrGlnPheLysValAspValProGlyAspTYrValLeu-----Va 332
537 GCAGGTCAAGGCGCTTACGCGCCAGAGACGCAACGCTTACTCTCCGTCGCGCACCGAT 596
Qy 332 AspHisAlaIle-----PheArgAlaPheAsnLysGlyAlaLeuGly----- 346
597 CGAACAATGATTTGCTCCAGCTCGTACGAATAGCGACAGGGCTGTGGTGTGCTGTC 656
Qy 347 -----IleLeuLysValGlyGlyGlnAsnHisGlyLys 360
657 TGAAGACGATTTGCGGATTCCTGAG--AGCGCCCTATACGACATTCGCGCTTC 713
Qy 360 HisLysGlnThrAspAlaVal---TYr--LeuProGlyLysAlaProGlnAlaIleAsp 379
714 CGGTGCAATGCTGACGTGTGTGCTTACGACCCAGCACTGAGACCGACGAGATTCGA 773
Qy 379 hrGlnGlnAla-----ProLysThrProAlaProAlaAsnLeu--G 392
```

```

Db      774 CGCAACG-GCCAACTGTAAGAGCATGCGCGCTCGGAGCAAAACCTGGGTC 832
Qy      392 InGUgInIleLy-----AlaGlyLysA 400
Db      833 AGGATGACGGCAAGCCAGCGCTGCTCGAAGCCGGTGCAAGGTATGACAGCGCAG 892
Qy      400 laThrTyraSPseRAsnCysAlaAlaCySHisGlnProAspGlyLysGlyValProAsnA 420
Db      893 AGGTTTACCTCCAACTGCTGCTCCATCTGCTATGAAGATGAGCATGCTGTCTCAACGCA 952
Qy      420 laPheProPLeuAlaAsnSerAspTyrlEuAsnAlaAspHisAlaArgAla--AlaS 439
Db      953 TGTTCGGCGCGCTGGTGGCAACCGGCTGTCATCAGCAAGCAACCTCAATGAGCCA 1012
Qy      439 erIleValAlaAsnGlyLeuSerGlyLysIleThrValAsnGlyAsnGlnTyrluSerV 459
Db      1013 ACATCGTACATTGGG-----GGTATTCTGCTCCGACGAATACGGCCGATCTGCTG 1066
Qy      459 aL---MetProAlaIleAla-----LeuSerAspGlnGlnIleAlaAsnValIleThrT 476
Db      1067 TTGCCATGCGCGGCTTCGCGCATCTGTCTGACACAGCATGCGCATGTTGTGAAC 1126
Qy      476 yThrIleuAsnSerPheGlyAsnLysGly---GlyGlnLeuSerAlaAspAspValAlaL 495
Db      1127 TCATGCCCAAGAGCTGGGGCAACAGGCTCGGAAACCTGTCTGCTCGGATATCCGCA 1186
Qy      495 ysaAlaLys 497
Db      1187 AGCTCCGC 1194

```

## RESULT 4

```

US-09-296-284-2
; Sequence 2, Application US/09296284A
; Patent No. 6204040
; GENERAL INFORMATION:
; APPLICANT: Choi, Eun-Sung
; APPLICANT: Rhee, Sang-Ki
; APPLICANT: Lee, Eun-Hae
; TITLE OR INVENTION: Gluconobacter Suboxydans Sorbitol Dehydrogenase, Genes
; FILE OR INVENTION: and Method of Use Thereof
; TITLE REFERENCE: 1533.0870000
; CURRENT APPLICATION NUMBER: US/09/296,284A
; CURRENT FILING DATE: 1999-04-22
; NUMBER OF SEQ ID NOS: 87
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 1437
; TYPE: DNA
; ORGANISM: Gluconobacter suboxydans
US-09-296-284-2

```

## Alignment Scores:

```

Pred. No.: 1,41e-07 Length: 1437
Score: 163.00 Matches: 118
Percent Similarity: 35.85% Conservative: 34
Best Local Similarity: 27.83% Mismatches: 144
Query Match: 6.21% Indels: 130
Gaps: 25

```

US-10-088-045-2 (1-502) x US-09-296-284-2 (1-1437)

```

Qy      171 LeuGlnProGlyLeuTyrrValTyrrHis-----CysAlaValAlaPro 184
Db      130 ATTAGGCGGGTGGCTTACGTGGCCCGCTGTCTACTGCGTGTGCTGCATACCGCACTA 189
Qy      185 ValGlyMetHisIleAlaAsnGlyMet-----TyrrGly 195
Db      190 CAGGCGCAGCTTTTGTCTGCTGCTGTGAGATCAAGAGCCGATCGGACAGCATCTACTCC 249
Qy      196 Leu---IleLeuValGluProLysGlnGlyLeuProLysValAspLysGluTyrrVal 214
Db      250 ACCAACATCAAGCTGAGCCGAAATACGGTATCGGCAAC-----TATACA 294

```

```

Qy      215 MetGlnGlyAspPheTyrrThrLysGlyLysTyrrGlyGlnGlnIleuGlnProPheAsp 234
Db      295 CTCGAAGAT-----TTCAGAAAGCG----- 315
Qy      235 MetGlnLysAlaIleArgGluAsp----- 242
Db      316 ATCCGTAAAGGTATCCGCAAGAGCAGCGCGACGGTTTATCCGCGCATGCCGTATCTGAG 375
Qy      243 -----AlaGluTyrrValValPheAsnGlySerVal 252
Db      376 TTGCTGCGCTGTCTGATGACGACATCAAGGCCATGATGCTTCTTCATGCAAGCGCTG 435
Qy      253 GlyAla-----LeuThrGlyGlnAsnAlaLeu 261
Db      436 AA-GCCGCTGCGCTTCAGAAACAGACCGGACATCTCTGCGCCATGAACATGCGCTG 494
Qy      262 LysAlaLysValGlyLthrValArg-LeuPheValGlyAsnGlyLysProAsnLeuTh 281
Db      495 -----GCCGTTGGCATCTGGCGCGCATGTTGTTCCGACTGTCAACACAGCGCTCGA 548
Qy      281 rSerSerPheHis-----ValIleGlyGlnIlePheAspLysValHisPheG 297
Db      549 CAAGACCATCTCCGATCCGAAAGTGGCGCGTGGCAATACCTC-----GTGAATGAGCC 602
Qy      297 uGlyGlyLysGlyLysAsnHisAsnIleGlnThrThrLeuIleProAlaGlyValAlaL 317
Db      603 AGGCCATTGGGCGCATGTCATACG-----CCCCGTGGCATGGCCAT 644
Qy      317 aIleThrGluPheLysValAspValProGlyAspTyrrValLeu-----Val 332
Db      645 GCAGTCAAGGGGTATACGGCCAAAGACGGCAACGCTTACTCTCCGTTGGCGCACCGAT 704
Qy      332 laPheHisAlaIle-----PheArgAlaPheAsnLysGlyAlaLeuGly----- 346
Db      705 CGCAACTGGATTGCTCCAGACCTGGTAGCAATACGACACGCGTGGTGGCTGGTC 764
Qy      347 -----IleuLysValGlnGlyGlnLysAsnHisGlnIleTyrrse 360
Db      765 TGAAGACGACATTGCCGAGTCTCTGAAG--AGCGCCGTTATGACATCTTCCGCTT 821
Qy      360 rHisLysGlnThrAspAlaVal---Tyrr-LeuProGlnGlyAlaProGlnAlaIleAspT 379
Db      822 CGGTGCATGCTGACGTGGTGGCTTACAGACCCAGCATGGACGACGACGATCTGCA 881
Qy      379 hrrGlnGluAla-----ProLysThrProAlaProAlaAsnLeu---G 392
Db      882 CGCAACG-GCCAACTGTAAGAGCATGCGCGCTTCCGAAAGCAAAACCTGAGTTC 940
Qy      392 InGUgInIleLy-----AlaGlyLysA 400
Db      941 AGGATGACGGCAAGGCCAGCGCCCTGCTCGAAGCCGCTGGCAAGGTATGACAGCGCAG 1000
Qy      400 laThrTyraSPseRAsnCysAlaAlaCySHisGlnProAspGlyLysGlyValProAsnA 420
Db      1001 AGGTTTACCTCCAACTGCTGCTCCATCTGCTATGAAGATGAGCATGCTGTGCAACGCA 1060
Qy      420 laPheProPLeuAlaAsnSerAspTyrlEuAsnAlaAspHisAlaArgAla--AlaS 439
Db      1061 TGTTCGGCGCGCTGGTGGCAACCGGCTGCTCATCAAGCAATGCAACCTCAATGAGCCA 1120
Qy      439 erIleValAlaAsnGlyLeuSerGlyLysIleThrValAsnGlyAsnGlnTyrluSerV 459
Db      1121 ACATCGTACATTGGG-----GGTATTCTGCTCCGACGAATACGGCCGCTGCTG 1174
Qy      459 aL---MetProAlaIleAla-----LeuSerAspGlnGlnIleAlaAsnValIleThrT 476
Db      1175 TTGCCATGCGCGGCTTCGCGCATCTGTCTGACACAGCATGCGCATGTTGTGAAC 1234
Qy      476 yThrIleuAsnSerPheGlyAsnLysGly---GlyGlnLeuSerAlaAspAspValAlaL 495
Db      1235 TCATGCCCAAGAGCTGGGGCAACAGGCTCCGGAACCTGTGCTCGGATATCCGCA 1294

```



QY 495 ysa1alys 497  
DB 1295 AGCTCCGC 1302  
RESULT 5  
US-09-296-284-7  
Sequence 7, Application US/09296284A  
Patent No. 6204040  
GENERAL INFORMATION:  
APPLICANT: Choi, Eun-Sung  
APPLICANT: Rhee, Sang-Ki  
APPLICANT: Lee, Eun-Hae  
TITLE OF INVENTION: Glucanobacter Suboxydans Sorbitol Dehydrogenase, Genes  
FILE REFERENCE: 1533.087000  
CURRENT APPLICATION NUMBER: US/09/296,284A  
NUMBER OF SEQ ID NOS: 87  
SOFTWARE: Patent In Ver. 2.0  
SEQ ID NO 7  
LENGTH: 4830  
TYPE: DNA  
ORGANISM: Glucanobacter suboxydans  
US-09-296-284-7  
Alignment Scores:  
Pred. No.: 9.37e-07 Length: 4830  
Score: 163.00 Matches: 146  
Percent Similarity: 34.40% Conservative: 187  
Best Local Similarity: 25.89% Mismatches: 187  
Query Match: 6.21% Indels: 186  
Gaps: 33  
US-10-088-045-2 (1-502) x US-09-296-284-7 (1-4830)  
QY 43 LysThrAlaAsn-----AlaAspAsnAlaAsnGlnGlnGlnGlnGlnLeu 59  
DB 2804 AAGCGCGCCAAAGCTCTACGACGAGAGTCAGAGCCGAAAGATGAGCGAGCGCTC 2863  
QY 60 ProValIleAspAlaIleValThrHisAlaProGluValPro-----ProPro 75  
DB 2864 CCTAACGGCGAATTCCTC-----AACGCTCCCTCAGGGTTCGTTGGCGCT 2908  
QY 76 ValAspArgAspHisProAlaValValValValMetGluThrValGluValMet 95  
DB 2909 GCAACCCCGGACATCCGTACCGGAAACCTC-----ACGCTGAAAGAAATGACG 2959  
QY 96 ArgLeuAlaAspGluValGluThrPheThrPheGlyGlyGlnVal-----112  
DB 2960 TGACATGCTCAAGGCAATTAC-----TCGGGACAGACTGGTATCTGA 3001  
QY 113 -----ProGlyGlnMetIleArgValArgGluGlyAspThrIleGluValGlnPheSer 130  
DB 3002 GATGAACAGGAGTGGAATACCGCGCGC-----3031  
QY 131 AsnHisProAspSerIleMetProHisAsnValAspPheHisAlaIleThrGlyProGly 150  
DB 3032 -----AGTCGGCTCATGAGAGTCTTTGGTGCTGCCAACCAGCA 3076  
QY 151 GlyGlyAlaGluAlaSerPheThrAlaProGlyHisThrSerThrPheSerPheVal 170  
DB 3077 C-GCTGATGACGCC-----CTG 3092  
QY 171 LeuGlnProGlyLeuThrValValValVal-----CysAlaValAlaPro 184  
DB 3093 ATTGACGGCGGCTCCATCGCGCCCTGTCTGACTGCGTTCCTCCATACCGGACTA 3152  
QY 185 ValGlyMetHisIleAlaGlnGlyMet-----TyrGly 195  
DB 3153 CAGGGCAGCGCTTTGCTGTGTCTGAGATCAAGACCCGATCGGACGATCTACTCC 3212  
QY 196 Leu---IleLeuValGluProGlyGluGlyLeuProGlyValAspIleGlyVal 214

DB 3213 ACCAATCATACCGCTGACCCGAAATACGGTATCGCAAC-----TATACA 3257  
QY 215 MetGlnGlyAspPheThrIleValGlyValGlyGlnGlnIleGlnProPheAsp 234  
DB 3258 CTCGAAGAT-----TTCAAGAGCGC-----3278  
QY 235 MetGluValAlaIleArgGluAsp-----242  
DB 3279 ATCCGTAAAGGTATCCGAAAGACCGCGGACGGTTATCCGGCATGCCGTATCTGAG 3338  
QY 243 -----AlaGluThrValValPheAsnGlySerVal 252  
DB 3339 TTCGCTCGCGCTGTGATGACATCAAGACCATGATATGCTTCTTCATGACGCGGT 3398  
QY 253 GlyAla-----LeuThrGlyGluAsnAlaLeu 261  
DB 3399 AA-GCCGCTCGCGCTTCAAGAACAGCCGACATCTCTGCGCATGAACTGGCGT 3457  
QY 262 LysAlaLysValGlyGluThrValArg-LeuPheValGlyAsnGlyGlyProAsnLeu 281  
DB 3458 -----GCCGTTGGCATCTGGCGCGCATGTTGTTCCGACTGTCAACACGACCTCGA 3511  
QY 281 rSerSerPheHis-----ValIleGlyGluIlePheAspLysValHisPheG 297  
DB 3512 CAAGACCATCTCCGATCCGAAAGTGGCGCTGGCGAAATACCTC-----GTGAATGCGCC 3565  
QY 297 uGlyGlyLysGlyGlyLysAsnHisAsnIleGlnThrThrLeuIleProAlaGlyValAla 317  
DB 3566 AGGCGCTTTGGCGAGTGTCTATCG-----CCCCGTGGCATGGCCAT 3607  
QY 317 aIleThrGluPheLysValAspValProGlyAspTyrValLeu-----Va 332  
DB 3608 GCAGCTCAAGGGGTATACGGCCAAAGACGGCAAGCTTACTCTCCGTTGGCGCACCAT 3667  
QY 332 LaspHisAlaIle-----PheArgAlaPheAsnLysGlyAlaLeuGly-----346  
DB 3668 CGACCACTGATGATCTCCAGCGCTGGCTAGCATACGACCGGCTGGTGGCTGCTGTC 3727  
QY 347 -----IleLeuLysValGluGlyGluGluGlnHisGluIleTyrSe 360  
DB 3728 TGACAGCATTTGCCGATTCCTGTAAG--ACGGCGGTATGACCATTTGCCGCTTT 3784  
QY 360 rHisLysGlnThrAspAlaVal--Tyr-LeuProGluGlyAlaProGlnAlaIleAsp 379  
DB 3785 CGGTGCGATGCGTACGCTGTGGCTTACAGCAACCGACGACCTGGACCGACGATCTGCA 3844  
QY 379 hGlnGluAla-----ProLysThrProAlaPheLeu--G 392  
DB 3845 CGCAACG-CCCAAGTACTGAAGACATGCCGCGCTCCGAAAGCAAAAACCTGGCTC 3903  
QY 392 IlnGlnIleLys-----AlaGlyLysA 400  
DB 3904 AGGATACCGGCAAGGCCACGCGCTGCTGAAACCGGTGGCAAGGTGATGACGGCGCAG 3963  
QY 400 lAthrTyrAspSerAsnCySAIAlaCySHIglrProAspGlyLysGlyValProAsnA 420  
DB 3964 AGGTTTACTCCACAACCTGCTGCTGCAATGACATGACATGACATGCTGCTCAACCGCA 4023  
QY 420 lAhpProProLeuAlaAsnSerAspTyrLeuAsnAlaAspHisAlaArgAla--AlS 439  
DB 4024 TGTTCCTCGCGCTGGCTGGCAACCGCGTGTATCATCGACGACATGCACTCACTGCGCA 4083  
QY 439 erIleValAlaAsnGlyLeuSerGlyLysIleThrValAsnGlyAsnGlnTyrGluSer 459  
DB 4084 ACATGCTGACATTGGC-----GGTATTCTGCTCCGACGAATACGGCGCATCTGCTG 4137  
QY 459 aI---MetProAlaIleAla-----LeuSerAspGlnGlnIleAlaAsnValIleThr 476  
DB 4138 TTGCATGCGCGGCTTCCCGCATCATCTGTGACCAAGACATCGCGCATGTTGTGAAC 4197  
QY 476 yThrThrLeuAsnSerPheGlyAsnLysGly--GlyGlnLeuSerAlaAspAspValAla 495  
DB 4198 TCATGCGCAAGAGCTGGGCGCAACGAGCTCCGGGAACCTGTCTGCTCGATATCCGA 4257



APPLICANT: FLEISCHMAN, Robert D.  
APPLICANT: WHITE, Owen R.  
APPLICANT: FRASER, Claire M.  
APPLICANT: VENTER, John C.  
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
FILE REFERENCE: 24366-2007.00  
CURRENT APPLICATION NUMBER: US/09/103,840A  
CURRENT FILING DATE: 1998-06-24  
NUMBER OF SEQ ID NOS: 2  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 1  
LENGTH: 4411529  
TYPE: DNA  
ORGANISM: Mycobacterium tuberculosis  
OTHER INFORMATION: H37RV  
US-09-103-840A-1

Alignment Scores:  
Pred. No.: 0.12 Length: 4411529  
Score: 158.50 Matches: 104  
Percent Similarity: 31.00% Conservative: 51  
Best Local Similarity: 20.80% Mismatches: 180  
Query Match: 6.04% Indels: 165  
Gaps: 23

US-10-088-045-2 (1-502) X US-09-103-840A-1 (1-4411529)

QY 16 SerAlaLeuMetLeuSerGlyCySerAsnGlnAlaAspLysAlaAlaGlnProLysSer 35  
DB 944111 ACGGCTTTCGCTACGCCCTTGC-----GCCTCGAAGCCACGACG 944070  
QY 36 SerThrValAlaAlaAlaLysThrAlaAsnAlaAspAsnAla-----AlaSerGln 54  
DB 944069 TTC-----GGGCCCCCGGAGATGACCCGTCGATGACGCGCGGAGGCGCGCG 944016  
QY 55 HisGlnGlyLeuLeuProValIleAspAlaIleValThrHisAlaProGluValProPro 74  
DB 944015 CACAGTGGCGA-----ACGGTTACCGCGACGCTGACC-----CCCCAG 943977  
QY 75 ProValAspArgAspHisProAlaLysValValValLysMetGluThrValGluLysVal 94  
DB 943976 CCGGCGAGATGACCTGGGTGGCGGATGTCAGCAGCCTG----- 943935  
QY 95 MetArgLeuAlaAspGlyValGluTyrGlnPheThrPheGlyGlyGlnValProGly 114  
DB 943934 -----ACCTACGGCAACACCATCCCGGA 943911  
QY 115 GlnMetIleArgValArgGluGlyAspThrIleGluValGlnPheSerAsnHisProAsp 134  
DB 943910 CCACTGATCCGGGCCACCGTGGGATGAGATTGCTGCTCGGTGACCAACCGTGGGT 943851  
QY 135 SerLysMetProHisAsnValAspPheHis-----AlaAlaThrGlyProGlyGly 152  
DB 943850 GAT-----CCGACGTCGGTGTGATGCGACGCGATGCGCTGGCGCAACGATGATGCG 943797  
QY 153 AlaGluAlaSerPheThr-----AlaProGlyHisThrSerThrPheSerPheLysAla 170  
DB 943796 ACCGAGCCCGGACGTCGGAACATCGGCCCGGGGTGACTTACCGGTTCTCGGT 943737  
QY 171 LeuGlnProGlyLeuTyrValTyrHisCysAlaValAlaProValGlyMetHisIleAla 190  
DB 943736 CCGGATCCGGGACCTACGCGGCCATCCGACGTCGCTTCACAGGCAAC----- 943683  
QY 191 AsnGlyMetTyrGlyLeuIleLeuValGluProLysGlyGlyLeuProLysValAspLys 210  
DB 943682 ---GGGCTATATCTGCTGCTGCTGCTGCGATCCACACTGAGCAGCAGCACTACGACGC 943626  
QY 211 GluTyrTyrValMetGlnGlyAspPheTyrThrLysGly----- 223  
DB 943625 GATGATGATCATCTCTCGACGAT---TGGACGAGACGCGATCGGGAAGTCCCGCAACAG 943569  
QY 224 LysTyrGlyGluGlnGlyLeuGlnProPheAspMetGluLysAlaIleArgGluAspAla 243

DB 943568 CTCTACGGGAGACTG-----ACCGACCCGCAACAAACCCATGCAAAACACA 943521  
QY 243 ----- 243  
DB 943520 ACAGTATCCCGAAGCGGAGCGTTGACACCAACTGTCGGCGGACGAGGGGAC 943461  
QY 244 -----GluTyrValValPheAsnGlySerValGlyAlaLeuThrGlyGluAsnAla 260  
DB 943460 ATCGCTACCGCTATCTGATGATCAACGGGCGCAATCCC-----GTGGCGGCACTCT 943407  
QY 261 LeuLysAlaLysValGlyGluThrValArgLeuPheValGlyAsnGlyGlyProAsnLeu 280  
DB 943406 TTAAAGGCCAACCTGGCCAGGAATCCGATCCGATCATGACACGCGCCGACACC 943347  
QY 281 ThrSerSerPheHisValIleGlyGluIlePheAspLysValHisPheGluGly----- 298  
DB 943346 GCGTTCGCGATCGCGTGGCGGCAATTGATGACGCGTACCCACACGACGTTACCCA 943287  
QY 299 -----GlyLysGlyLysAsnHisAsnIleGln 307  
DB 943286 GTGATTCACCGCAAGTGCAGCCTGCTGATCGCATGCGCAACGCTACGATCAG 943227  
QY 308 ThrThrLeuIleProAlaGlyGly-----AlaAlaIleThrGluPheLysVal 323  
DB 943226 GTGACC-----GCCGTGCGCGCGCTTCCCTCGTGCACCTCGCGAAGGCAAG----- 943176  
QY 324 AspValProGlyAspTyrValLeuValAspHisAlaIlePheArgAlaPheAsnLysGly 343  
DB 943175 -----AACGGCGCGCGCGTGGCTGCTGTACC 943146  
QY 344 AlaLeuGlyIleLeuLysValGluGlyGluGluAsnHisGluIleTyrSerHisLysGln 363  
DB 943145 GCGCGCGCG-----ACCCA 943131  
QY 364 ThrAspAlaValTyrLeuProGluGlyAlaProGlnAlaIleAspThrGlnGluAlaPro 383  
DB 943130 CCCGACCGCGATTCGCGCGGATGAACTCACTGCGGAGTGATGCGTGAATGTTCC 943071  
QY 384 LysThrProAlaProAlaAsnLeuGlnGlnIleLysAlaGlyLysAlaThrTyrAsp 403  
DB 943070 ACCGCGCAACGACTGCCAACCTG----- 943047  
QY 404 SerAsnCysAlaAlaCysHisGlnProAspGlyLysGlyValProAsnAlaPheProPro 423  
DB 943046 -----GGCGGCGCGCAACCCACCAACGACCTCCCG 943017  
QY 424 LeuAlaAsnSerAspTyrLeuAsnAlaAspHisAlaArgAlaAlaSerIleValAlaAsn 443  
DB 943016 -----CTCACCTTGGGC 943005  
QY 444 GlyLeuSerGlyLysIle-----ThrValAsnGlyAsnGlnTyrGluSerValMetPro 461  
DB 943004 GGCACCATGGCCAACTACGACTGCAATCAACGGGGAACCTTACGACGACCAATCCA 942945

RESULT 8  
US-09-252-991A-6541/C  
Sequence 6541, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubinfeld et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 6541  
LENGTH: 1302

TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-6541

## Alignment Scores:

Pred. No.:	8,466-07	Length:	1302
Score:	155.00	Matches:	87
Percent Similarity:	37.43%	Conservative:	38
Best Local Similarity:	26.05%	Mismatches:	109
Query Match:	5.90%	Indels:	100
DB:	4	Gaps:	16

US-10-088-045-2 (1-502) x US-09-252-991A-6541 (1-1302)

```

QY 115 GlnMetIleArgValAsgGluGlyAspThrIle-----GluValGlnPheSerAsnHis 132
DB 1216 AAGCTTTAGAGTAAAGAAATGACGACCTTTGAGCTGTGACATTACAGCTTCAC 1157
QY 133 ProAspSerLysMetProHisAsnValAspPheHisAlaAlaThrGlyProGlyGly 152
DB 1156 CCT-----CGGACCTGCTGGCGGCTGACCTGCTGATCTTCCGCAACCGCAAGGCCA 1103
QY 153 AlaGlu--AlaSerPheThrAlaProGlyHisThrSerThrPheSerPheLysAlaLeu 171
DB 1102 GCGCAGCAGACACACGACGACGACGACCTTCTACGACGACGACGACGACGACGACGAC 1043
QY 172 GlnProGlyLeuThrValTyrHisCysAlaValAlaProValGlyMetHis-----Ile 189
DB 1042 CAACCCCTGCTCCGAAGTG-----GTGCTTATCTGTTCTGCTGCGACCTGCTGCTT 992
QY 190 AlaAsnGlyMetTyrGlyLeuIle-----LeuValGlnProLysGly 204
DB 991 CGCGGTGGGCTACTGCGCCCTGTACCGGGCGCTGGACCTGGAAGGCTGATGCCGG 932
QY 205 LeuPro-LysValAlaPheLysGluTyrTyrValMetGlnGlyAspPheTyrThrLysGly 224
DB 931 CTACCACTCCGCCAC-----GAATTCGCCGCAAGGAAAA 896
QY 224 GTYrGlyGluGlnGlyLeuGlnProPheAspMetGluLysAlaIleArgLysAlaGly 244
DB 895 A--GGCTGACCGGCTCCACCACTGGAAGAAAGAAATGACC--AAGGCCAGCAGAA 842
QY 244 uTYrValValAlaPheAsnGlySerValGlyAlaLeuThrGlyGluAsnAlaLeuLys 262
DB 841 ATACGCCCGCATCTTCCCAAGTTCGCGCGGATCCCATGAGAAATCCGCAAGATCC 782
QY 263 -----AlaLysValAlaGlyLeuThrValAlaGluPheValGlyAsn----- 275
DB 781 GCAGGCGGTGAATAATGGCGGT-----CGCTGCTTCCCTCCAACTGCTCATCTCCA 728
QY 276 -----GlyGlyProAsnLeuThrSerPheHisValIle 287
DB 727 CGGCTCCGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 668
QY 287 egiLysIlePheAspLysValHisPheGluGlyGlyLysGlyGluAsnHisAsnIleGly 307
DB 667 GGGC-----GGCGAGCGGAAACCATCA 644
QY 307 nTYrThrLeuIleProAlaGlyAlaAlaIleThrGluPheLysValAspValProGly 327
DB 643 GACCAACCATCATG--GCTGGC----- 625
QY 327 yAspTyrValIleValAspHisAlaIlePheArgAlaPheAsnLysGlyAlaLeuGlyIle 347
DB 624 -----CGCCAGCGCGCATGCGGCTGGGGTGA-----GT 593
QY 347 eLeuLysValAlaGluGlyGluAsnHisGluLysSerHisLysGlnThrAspAlaVala 367
DB 592 GATCGCGGAGGAAAGCGTGAAGAAACGTCGCGCTTGTCTCTCCACCAAGATGATGGCCG 533
QY 367 lTYrLeuProGluGlyAlaProGlnAlaIleAspThrGlnGluAlaProLysThrProAl 387
DB 532 TAAGCTGCGGAAAGCGCC----- 514

```

```

QY 387 aProAlaAsnLeuGlnGluIleLysAlaGlyLysAlaThrTyrAspSerAsnCysAl 407
DB 513 -----AAGCAGACATTCAGGCCCGGCAAGCAGCTTGTGCTACACCTGCG 467
QY 407 aAlaCysHisGlnProAspGlyLysGlyValPro 418
DB 466 CGCCTGCCACGCTCCGAAAGGCAAGGACCCCG 433

```

## RESULT 9

US-09-252-991A-6490  
Sequence 6490, Application US/09252991A

```

; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 6490
; LENGTH: 963
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-6490

```

## Alignment Scores:

Pred. No.:	5,966-07	Length:	963
Score:	154.50	Matches:	87
Percent Similarity:	37.20%	Conservative:	35
Best Local Similarity:	26.52%	Mismatches:	107
Query Match:	5.88%	Indels:	99
DB:	4	Gaps:	16

US-10-088-045-2 (1-502) x US-09-252-991A-6490 (1-963)

```

QY 120 ArgGluLysAspThrIleGlu--ValGlnPheSerAsnHisProAspSerLysMetPro 138
DB 3 AAGATGACGACCTTGTGAGCTGTGATACATTACAGCTTCACCT-----CGGACCCCT 56
QY 139 HisAsnValAspPheHisAlaAlaThrGlyProGlyGlyAlaGlu--AlaSerPheT 158
DB 57 GCTGGCGCTGACCTGCTGATCTTCCGCAACCGCAAGGCGCAGCAGCAGCAGCAGCAG 116
QY 158 hTYrAlaProGlyHisThrSerThrPheSerPheLysAlaLeu-GlnProGlyLeuTyrVal 177
DB 117 CGAGACCGTGGGCACTCTTACGACGACGACGACGACGACGACGACGACGACGACGAC 176
QY 178 TyrHisCysAlaValAlaProValGlyMetHis-----IleAlaAsnGlyMetTyrGly 195
DB 177 -----GTGCTTATCTGCTTCTGCGGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 227
QY 196 LeuIle-----LeuValGlnProLysGlyLysLeuPro-LysValAspLys 210
DB 228 CCGTACCGCGGCTGGGCACTGGAAGGCTGATCCGGGCTTACAGTCCGCCGAC-- 285
QY 210 sgiLysTyrValMetGlnGlyAspPheTyrThrLysGlyLysTyrGlyGluGlnGlyLys 230
DB 286 -----GAATTCGCCGCAAGGAAAAA--GCGTGAACCGGCT 320
QY 230 uGlnProPheAspMetGluLysAlaIleArgLysAlaGluTyrValValAlaPheAsnGly 250
DB 321 CCACAGTGTGGAAGAAATGAGCC--AAGCGCAGCAAGAAATACGCGCCGATCTTCCG 377
QY 250 ySerValGlyAlaLeuThrGlyGlyLysAlaLeuLys-----AlaLysValGly 266
DB 378 CAAGTTCGCGCGATCCATGAGAAAGTCCGCAAGGATCCGACGCGGTGAATAATGGG 437

```

QY 266 ygluThrValArgLeuPheValGlyAsn----- 275  
DB 438 CGGT-----CGCTGTTCCCTCCCACTGCTCATCTGCCAGCGCTCCGACCCCAAGG 491  
QY 276 -----GlyGlyProAsnLeuThrSerSerPheHisValIleGlyGluLeuPheAsp 293  
DB 492 CGCTTACGGGTTTCCCACTGACCGACGCGCATGCGCTGGGGC----- 537  
QY 293 sValHisPheGluGlyGlyLeuSerHisAsnIleGlnThrThrLeuLeuProAl 313  
DB 538 -----GGCAGCGCGGAAACCATTAAGACCACTCATG---GC 572  
QY 313 aglyGlyAlaAlaIleThrGluPheValAspValProGlyAspTyrValLeuValAs 333  
DB 573 TGGC-----CG 578  
QY 333 pHisAlaIlePheArgAlaPheAsnIleGlyAlaLeuGlyIleLeuLysValGluGly 353  
DB 579 CCAGCGCGCGCATGCCGCTGGGGTGA-----GTGATCGCGCAGGAGAGCGCT 626  
QY 353 uGluAsnHisGluIleTyrSerHisLysGlnThrAspAlaValTyrLeuProGluGlyAl 373  
DB 627 GAAGAACGTCGCGCTTCTGCTCCTCACCCAGATGATGCGCTGAAGTCCGGAAGCGC 666  
QY 373 aProGlnAlaIleAspThrGlnGluAlaProLysThrProAlaProAlaAsnLeuGlnI 393  
DB 687 C-----AAGC 692  
QY 393 uGlnIleLysAlaGlyLysAlaThrTyrAspSerAsnCySAAlaCySHisGlnProAs 413  
DB 693 AGACATCGAGGCGCGGAGAGGCTTCTGCTACACCTGCTGCTGCGCAGCGTCCGGA 752  
QY 413 pGlyLysGlyValPro 418  
DB 753 AGGCAAGGCGACCCCG 768

## RESULT 10

US-09-252-991A-9169/c  
; Sequence 9169, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252.991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074.788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094.190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 9169  
; LENGTH: 1470  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1314)  
; OTHER INFORMATION: Identity of nucleotide at the above locations are unknown.  
US-09-252-991A-9169

## Alignment Scores:

Pred. No.: 1.67e-06 Length: 1470  
Score: 153.00 Matches: 38  
Percent Similarity: 49.53% Conservative: 15  
Best Local Similarity: 35.51% Mismatches: 48  
Query Match: 5.83% Indels: 6  
DB: 4 Gaps: 4

US-10-088-045-2 (1-502) x US-09-252-991A-9169 (1-1470)

QY 396 LysAlaGlyLysAlaThrTyrAspSerAsnCySAAlaCySHisGlnProAspGlyLys 415

DB 453 AAGACCGCGCGCGCTTACCTGACCAACTCGGGGCTGCGACCGCACCGCGCAAG 394  
QY 416 GlyValProAsnAlaPheProProLeuAla---AsnSerAspTyrLeuAsnAlaAspHis 434  
DB 393 GGCTATCGCGGGTCTTCCCGGCTGCGCGGCAACCGGTATGACCGGCGAGCCCG 334  
QY 435 AlaArgAlaAlaSerIleValAlaAsnGlyLeuSerGlyLysIleThrValAsnGlyAsn 454  
DB 333 ACTCCCTGGTGACATCGTGTCAAGGCG-----GGCACCCTGCGGCGCACCCACG 280  
QY 455 GlnTyrGluSer---ValMetProAlaIleAla-----LeuSerAspGlnGlnIleAla 471  
DB 279 GCACCTCGAGCTTCACCATGCGCGCTGCGCTGCGGATGACACACGAGATCGGC 220  
QY 472 AsnValIleThrTyrThrLeuAsnSerPheGlyAsnIleGlyGlyGlnLeuSerAlaAsp 491  
DB 219 GAGTGTCACTCTCATCCGACCACTGGGCGACAGGCGCAAGCGTCAAGCTCGAC 160  
QY 492 AspValAlaLysAlaLysLys 498  
DB 159 GAGTCAAGCGCGCTGCGCAAG 139

## RESULT 11

US-09-252-991A-8797  
; Sequence 8797, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252.991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074.788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094.190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 8797  
; LENGTH: 1764  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-8797

## Alignment Scores:

Pred. No.: 2.22e-06 Length: 1764  
Score: 153.00 Matches: 38  
Percent Similarity: 49.53% Conservative: 15  
Best Local Similarity: 35.51% Mismatches: 48  
Query Match: 5.83% Indels: 6  
DB: 4 Gaps: 4

US-10-088-045-2 (1-502) x US-09-252-991A-8797 (1-1764)

QY 396 LysAlaGlyLysAlaThrTyrAspSerAsnCySAAlaCySHisGlnProAspGlyLys 415  
DB 845 AAGACCGCGCGCGCTTACCTGACCAACTCGGGGCTGCGACCGCACCGCGCAAG 904  
QY 416 GlyValProAsnAlaPheProProLeuAla---AsnSerAspTyrLeuAsnAlaAspHis 434  
DB 905 GGCTATCGCGGGTCTTCCCGGCTGCGCGGCAACCGGTATGACCGGCGAGCCCG 964  
QY 435 AlaArgAlaAlaSerIleValAlaAsnGlyLeuSerGlyLysIleThrValAsnGlyAsn 454  
DB 965 ACCCTCGGTGTCACATCGTGTCAAGGCG-----GGCACCCTGCGGCGCACCCACG 1018  
QY 455 GlnTyrGluSer---ValMetProAlaIleAla-----LeuSerAspGlnGlnIleAla 471  
DB 1019 GCACCTCGAGCTTCACCATGCGCGCTGCGCTGCGGATGACACGAGATCGGC 1078  
QY 472 AsnValIleThrTyrThrLeuAsnSerPheGlyAsnIleGlyGlyGlnLeuSerAlaAsp 491

Db 1079 GAGGTGTCATCTTCACGACGCTGGGCAACGAGCGGCGTACGCTGAC 1138  
Qy 492 AspValAlaLysAlaLysLys 498  
Db 1139 GAGCTCAGCGCGCTGCGCAAG 1159

## RESULT 12

US-09-252-991A-8910  
Sequence 8910, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252.991A  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 8910  
LENGTH: 2256  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (1015)  
OTHER INFORMATION: Identity of nucleotide at the above locations are unknown.  
US-09-252-991A-8910

## Alignment Scores:

Pred. No.:	3,26e-06	Length:	2256
Score:	153.00	Matches:	38
Percent Similarity:	49.53%	Conservative:	15
Best Local Similarity:	35.51%	Mismatches:	48
Query Match:	5.83%	Indels:	6
DB:	4	Gaps:	4

US-10-088-045-2 (1-502) x US-09-252-991A-8910 (1-2256)

Qy 396 LysAlaGlyLysAlaThrTrpAspSerAsnGlyAlaIleCysHisGlnProAspGlyLys 415  
Db 1876 AAGACGCGCGCGGCTCTACGTGACACATGCGGCGCGCCACCGACGCGCAAG 1935  
Qy 416 GlyAlaProAsnAlaPheProProLeuAlaLysSerAspTrpLeuAlaAspHis 434  
Db 1936 GGCTTATGCGCGGCTCTTCCCGCCCTGCGCAACCGGTAGTACCGGCGGACCGCG 1995  
Qy 435 AlaArgAlaAlaSerIleValAlaAsnGlyLeuSerGlyLysIleThrValAlaAsnGlyAsn 454  
Db 1996 ACCCTCCTGCTGACATCGTCTCAAGGC-----GACACCTCGCGCGGCGGCGGACCG 2049  
Qy 455 GlnTrpGluSer---ValMetProAlaIleAla-----LeuSerAspGlnGlnIleAla 471  
Db 2050 GCACCGTCGAGCTTCACCATGCGCGCTGCGGTGGCGATGAACGACGAGATCGCC 2109  
Qy 472 AsnValIleThrTrpThrLeuAsnSerPheGlyAsnLysGlyGlnLeuSerAlaAsp 491  
Db 2110 GAGCTGTCACCTATCATCCGACGAGCTGGGCAACGAGCGCGCCAGCTCAGCGTGCAC 2169  
Qy 492 AspValAlaLysAlaLysLys 498  
Db 2170 GAGCTCAGCGCGCTGCGCAAG 2190

## RESULT 13

US-07-985-458-2  
Sequence 2, Application US/07985458  
Patent No. 5344777  
GENERAL INFORMATION:  
APPLICANT: Tamaki, Toshimi;  
APPLICANT: Takemura, Hiroshi;

APPLICANT: Tayama, Kenji;  
APPLICANT: Fukaya, Masahiro;  
APPLICANT: Okumura, Hajime and  
APPLICANT: Kawamura, Yoshiya  
TITLE OF INVENTION: Structural Gene of Membrane-Bound  
TITLE OF INVENTION: Alcohol Dehydrogenase Complex, Plasmid  
TITLE OF INVENTION: Containing The Same And Transformed Acetic Acid  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Frischauf, Holtz, Goodman & Woodward, P.C.  
STREET: 600 Third Avenue  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10016-2088  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 Inch, 0.72 mb  
COMPUTER: IBM PC compatible (NEC PC-9801 ES)  
OPERATING SYSTEM: MS DOS  
SOFTWARE: ASCII Form  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/985,458  
FILING DATE: 19921203  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/658,221  
FILING DATE: 20-FEB-1991  
APPLICATION NUMBER: 73440/1990  
FILING DATE: 26-MAR-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Goodman, Herbert  
REGISTRATION NUMBER: 17061  
REFERENCE/DOCKET NUMBER: 910134/HG  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212)972-1400  
TELEFAX: (212)370-1622  
TELEX: 236268  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1404 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: unknown  
MOLECULE TYPE: genomic DNA  
ORIGINAL SOURCE:  
ORGANISM: Acetobacter altoacetigenes  
STRAIN: MH-24  
PUBLICATION INFORMATION:  
AUTHORS: Tamaki, Toshimi;  
AUTHORS: Fukaya, Masahiro;  
AUTHORS: Takemura, Hiroshi;  
AUTHORS: Tayama, Kenji;  
AUTHORS: Okumura, Hajime;  
AUTHORS: Kawamura, Yoshiya;  
AUTHORS: Nishiyama, Makoto;  
AUTHORS: Horinouchi, Sueharu and  
AUTHORS: Beppu, Teruhiko  
TITLE: Cloning and Sequencing of the Gene Cluster  
TITLE: Encoding Two Subunits of Membrane-Bound  
TITLE: Alcohol Dehydrogenase from Acetobacter  
TITLE: polyoxogenes  
JOURNAL: Biochimica et Biophysica Acta.  
VOLUME: 1088  
PAGES: 232-300  
DATE: 1991

## US-07-985-458-2

Alignment Scores:  
Pred. No.: 1.98e-06  
Score: 152.00  
Percent Similarity: 38.02%  
Best Local Similarity: 24.88%

Length:	1404
Matches:	108
Conservative:	57
Mismatches:	165

Query Match: 5.79% Indels: 104  
 DB: 1 Gaps: 23  
 US-10-088-045-2 (1-502) x US-07-985-458-2 (1-1404)

QY 113 ProGlyGlmethleatrgValarg-----GluGlyAspThrIleGluValGln 128  
 DB 156 CCGTGGACAGCTCGTACGCGCGGCTTGAATCAAGAGCCGATCGTACGATCTACTC 215  
 QY 129 PheSerAsnHis-----ProAsp----- 134  
 DB 216 CACGAACATCACACCGGACCGGACCTTACGGTATCGCTTACACTTCCCGCAATTGCA 275  
 QY 135 ---SerIleMetPro-----HisAsnValAspPheHisAlaAla 146  
 DB 276 CGAAGCGTGCCTGATCGGCAAGACGCGTCCAGCGCTGATCCGCGCATGCCGTA 335  
 QY 147 Thr-----GlyProGlyGlyValaGluAlaSerPheThrAla 159  
 DB 336 TCCCTCTCTTCCGCGATGACGAAGAGACATGCGCGCTGTATGCGTACTTCATGCA 395  
 QY 160 ProGlyHisThrSerPheSerPheValAlaLeuGlnProGlyLeuValTyrHis 179  
 DB 396 TGGGGTGAAGCCGCTGCGCGACGCGACAGACGACATCTCTGCGCTTGTTCAT 455  
 QY 180 CysAlaValAlaProValGlyMetHisIleAlaAsnGlyMetTyrGlyLeuIleVal 199  
 DB 456 GCGCTGGCGCTGG-----CATCTGCGCATCAT-----GTTCTCGCC 494  
 QY 200 -GluProGlyGlyLeuProGlyValaAspGlyTyrTyrValMetGlnGlyAsp 219  
 DB 495 TTGCGCGAAGAGCTTACGCGCGCGCGACGAGCTTGAATCGACGATGCGCGATTA 554  
 QY 219 eTyrThrIleGlyTyrGlyTyrGlyGlnGlyLeu-----GlnProPheAsp 235  
 DB 555 TCTGTTACCGGCGCC---GGGCAATTGCGGTGCTGATACCCCGGTGGCTTCCGAT 611  
 QY 235 ---GluTyrAlaIleArgGluAspAlaGluTyrValAlaPheAsnGlySerVal-GlyA 254  
 DB 612 GCAGGAAAGCGCTGGACGCTCGCGTGTCTGACTTCTGTGCGTGGCGCACCGAT 671  
 QY 254 ILeuThrGlyGluAsnAlaLeuAlaValaGlyGlnThrValaArgLeuPheValG 274  
 DB 672 CGACACTGGGTGCGCGCGCGCGCTGCGCA-----ACGATCTGTGCTGTGCT 719  
 QY 274 IAsnGlyProAsnLeuThrSer---SerPheHisValIleGlyGluIlePheAsp 293  
 DB 720 GGGCGCGCTGTCGAGATGACATCTACACCTTCTGAACTCGCGCGGTTCGACCATC 779  
 QY 293 sValHisPheGluGlyGlyTyrGlyGluAsnHisAsnIleGlnThrThrLeuIleProAl 313  
 DB 780 CCGCGCTTTC---GGTGGCATGGGCGAT----- 804  
 QY 313 aGlyGlyAlaAlaIleThrGluPheValaAspValaProGlyAspTyrValaLeuVala 333  
 DB 805 ---GTGGTGCATGAGACACCCAGTACTTACCGCATGACGACCTG----- 846  
 QY 333 pHisAlaIlePheArgAlaPheAsnGlyAlaLeuGlyIleLeuValaGluGly 353  
 DB 847 -CACGCGCATCGCAGAG---TACCTGAAGAGCTGCGCGCGCGCTGCGCGCAAGGCA 902  
 QY 353 uGluAsnHisGluIleTyrSerHisGlyGlnThrAspAlaValaTyrLeuProGluGlyAl 373  
 DB 903 C-----TACACCTACGATCCGCTCACCGCAACATGCTGCTGCGGT-- 945  
 QY 373 aProGlnAlaIleAspThrGlnGluAlaProGlyThrProAlaProAlaAsnLeuGlnI 393  
 DB 946 ---AATACCGCCAGCGCTTCG----- 963  
 QY 393 uGlnIleValaGlyTyrValaThrTyrAspSerAsnGlyAlaAlaGlyHisGlnProAl 413  
 DB 964 -----GGTGTGATACGTATGTGAAGAAAGGCGCATCTGTACCGCTAAGCA 1010

QY 413 pGlyTyrGlyValaProAsnAlaPheProLeuAlaAsnSerAspTyrLeuAsnAlaAs 433  
 DB 1011 CCGTGGTGGCGTGGCGCGCATGTCTCCGCGCTGCGCAACCCGCTGTCTGACCGCA 1070  
 QY 433 pHisAlaAlaGala---AlaSerIleValaAlaAsnGlyLeuSerGlyTyrIleThrVala 452  
 DB 1071 GAACCCGACCTCGCTGTGTAACGTATGCGCATGTGT-----GGGTGCTGCGCGCAG 1124  
 QY 452 nGlyAsnGlnTyrGluSerVal---MetProAlaIle-----AlaLeuSerAspGlnI 469  
 DB 1125 CAACCTGGGACCGCTCCGCGATGCAATGCGGCTTACAGCACTGCTGCGCGCACGA 1184  
 QY 469 nIleAlaAsnValaIleThrTyrThrLeuAsnSerPheGlyAsnGlyTyr---GlyLeu 488  
 DB 1185 GATTGCTGATGTGTCACTTATCCGACACGACTGGGCGCAACAGCGCGCACCGCT 1244  
 QY 488 uSerAlaAspAspValaAlaAlaValaValaValaValaValaValaValaValaVala 499  
 DB 1245 TACGCTGCGGATGTACCAAGCTGCGCACAGC 1278

RESULT 14  
 US-09-252-991A-765  
 ; Sequence 765, Application US/09252991A  
 ; Patent No. 6551795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: 107196.136  
 ; CURRENT APPLICATION NUMBER: US/09/252, 991A  
 ; CURRENT FILING DATE: 1999-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/074, 788  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/094, 190  
 ; PRIOR FILING DATE: 1998-07-27  
 ; NUMBER OF SEQ ID NOS: 33142  
 ; SEQ ID NO 765  
 ; LENGTH: 1335  
 ; TYPE: DNA  
 ; ORGANISM: Pseudomonas aeruginosa  
 ; US-09-252-991A-765

Alignment Scores:  
 Pred. No.: 2,066-06 Length: 1335  
 Score: 151.50 Matches: 87  
 Percent Similarity: 31.83% Conservative: 40  
 Best Local Similarity: 21.80% Mismatches: 110  
 Query Match: 5.77% Indels: 162  
 DB: 4 Gaps: 18

US-10-088-045-2 (1-502) x US-09-252-991A-765 (1-1335)

QY 132 HisProAspSerIleMetProHisAsnValaAspPheHisAlaIleThrGlyProGlyGly 151  
 DB 528 CACCTGTAT-----CCACATCTACGACACTTCCGAGCCGAGCTGACCTGACGAT 578  
 QY 152 ---GlyAlaGluAlaSerPheThrAlaProGlyHisThrSerPhe 168  
 DB 579 CACCGGCTACCACTGGAAGTGCATGACAGTACCTGG----- 617  
 QY 169 LysAlaLeuGlnProGlyLeuTyrValaTyrHisCysAlaValaAlaProValaGlyMetHis 188  
 DB 618 -----CCAGAGCTGAGTACCTTACGCAACCTGCGCACCGCGCAGGA----- 659  
 QY 189 IleAlaAsnGlyMetTyrGlyLeuIleLeuValaGluProGlyGluGlyLeuPro-LysVal 208  
 DB 660 ---CCAGATCCACAA-----CCGACGACGC 680  
 QY 208 IAspGlyTyrTyrValaMetGlnGlyAspPheTyrThrTyrGlyTyrGlyGlu 228  
 DB 681 GAAGACGAGCATTACTCTGGAGGTGAC----- 711  
 QY 228 nGlyLeuGlnProPheAspMetGluValaIleArgGluAspAlaGluTyrValaValaPh 248

```

Db      711 ----- 711
Qy      248 eAenGlySerValGlyAlaLeuThrGlyGluAsnAlaLeuValAlaLeuValGlyLuth 268
Db      712 ----- 740
Qy      268 rValArgLeuPheValGlyAenGlyGlyProAsnLeuThrSerSerPheHisValIleg1 288
Db      741 GGTGGCTTCTGATC-----ACCTCCAGC----- 765
Qy      288 yGluIlePheAspLysValHisPheGluGlyLysGlyValAsnHisValIleg1Lth 308
Db      766 -----GACGTGATCCAT-----TC 779
Qy      308 rThrLeuIleProAlaGlyAla-----AlaIleThrGluPhe----- 321
Db      780 CTGTGGTGTCCCGCTTCCGGGTCAAGCCGAGCCATCCCGGCTTCTCAACAGAGC 839
Qy      322 -----LysValAspValProGlyAspTyr----- 329
Db      840 CTGACCAAGGTCCAGCCGCGCATCTATCGCGCAAGTGGCCGAGGTGCGGCA 899
Qy      330 -----ValLeuValAspHisAlaIlePheArgAlaPheAsnLysG1 343
Db      900 GACACGAGCTTCATGCCGATCGTGTGAC----- 930
Qy      343 yAlaLeuGlyIleLeuLysValGlyGlyGluAsnHisGluIleTyrSerHisLysG1 363
Db      931 -----GTCAAGCCCAAGCCGAGTTCGACAGTGGCTGGCCCAAGCCAGGA 977
Qy      363 nThrAspAlaValTyrLeuProGluGlyAlaProGluAlaIleAspThrGluAlaPyr 383
Db      978 A---GAGCGCGCAAGGTCAAGAA-----CTGACCAAGCAAGAGTGGAC 1019
Qy      383 oLysThrProAlaProAlaAsnLeuGluGlnIleLysValGlyLysAlaThrTyrAs 403
Db      1020 CAAG-----GAAAGTGTGGTGGCGGCGGCAAGGTCTACCA 1058
Qy      403 pSerAsnGlyAlaAlaCysHisGlnProAspGlyLysGlyValProAsnAlaPheProPyr 423
Db      1059 CACCATTCGGCCCGCTTCGCCACAGCCGAGGCGCAAGGCGCATGTTCCCGGC 1118
Qy      423 oLeuAlaAsnSerAspTyrLeuAsnAlaAspHisAlaArgAlaAlaSerIleValAlaAs 443
Db      1119 GCTGAAGGTTTCAAGATCGTCAAGCGGCCCAAGAGACACCACTGGAAGTCTTCA 1178
Qy      443 nGlyLeuSerGly---LysIleThrValAsnGlyAsnGlnTyrGluSerValMetProAl 462
Db      1179 CGGCGTCCCGGCGCACCGCATGGCGGCTTCGGCAAGCAG----- 1218
Qy      462 aIleAlaLeuSerAspGlnGlnIleAlaAsnValIleThrTyrThrLeuAsnSerPheG1 482
Db      1219 -----CTCAACGAGTCAACCTGGCGCGGTGATACCTTACAGCGCAACGCTGGGG 1271
Qy      482 yAsnLysGlyGlyGlnLeuSerAlaAspValAlaLysAlaLysLysThrLys 500
Db      1272 CAACGACGATGGGACATGTCACCCGAAAGACGTGTGCTCAAGACAGAA 1326

```

```

RESULT 15
US-09-252-991A-737
/ Sequence 737, Application US/09252991A
/ Patent No. 6551795
/ GENERAL INFORMATION:
/ APPLICANT: MARC J. RUBENFELD et al.
/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
/ FILE REFERENCE: AERGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
/ CURRENT APPLICATION NUMBER: US/09/252,991A
/ PRIOR FILING DATE: 1998-02-18
/ PRIOR APPLICATION NUMBER: US 60/074,788
/ PRIOR FILING DATE: 1998-02-18
/ PRIOR APPLICATION NUMBER: US 60/094,190

```

```

/ PRIOR FILING DATE: 1998-07-27
/ NUMBER OF SEQ ID NOS: 33142
/ SEQ ID NO 737
/ LENGTH: 1524
/ TYPE: DNA
/ ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-737

Alignment Scores:
Pred. No.: 2,54e-06 Length: 1524
Score: 151.50 Matches: 87
Percent Similarity: 31.83% Conservative: 40
Best Local Similarity: 21.80% Mismatches: 110
Query Match: 5.77% Indels: 162
DB: Gaps: 18

US-10-088-045-2 (1-502) x US-09-252-991A-737 (1-1524)

Qy      132 HisProAspSerLysMetProHisAsnValAspPheHisAlaIleThrGlyProGlyGly 151
Db      436 CACCTGAT-----CCACATCTACGACACTTCCAGCCGAGCTGGACGTGACGAGT 486
Qy      152 -----GlyAlaGlyAlaSerPheThrAlaProGlyHisThrSerThrPheSerPhe 168
Db      487 CACCGGCTACCACTGGAAAGTGCAGTACAGTACCTGG----- 525
Qy      169 LysAlaLeuGlnProGlyLeuTyrValTyrHisCysAlaValAlaProValGlyMetHis 188
Db      526 -----CCAGAGCTGAGTACTTTCAGCAACCTGGCCAGCCCGAGGA----- 567
Qy      189 IleAlaAsnGlyMetTyrGlyLeuIleLeuValGlnProLysGluGlyLeuPro-LysVa 208
Db      568 -----CCAGATCCACA-----CCGAGCAGGC 588
Qy      208 LAspLysGlnTyrTyrValMetGlnLysPheTyrThrLysGlyLysTyrGlyGluG1 228
Db      589 GAAGACGAGCATACCTCTGGAGGTGAC----- 619
Qy      228 nGlyLeuGlnProPheAspMetGluLysAlaIleArgGluAspAlaGluTyrValAlaPh 248
Db      619 ----- 619
Qy      248 eAenGlySerValGlyAlaLeuThrGlyGluAsnAlaLeuValAlaLeuValGlyLuth 268
Db      620 -----GAGCCGCTGGTCCGCGTGGCGACCA 648
Qy      268 rValArgLeuPheValGlyAenGlyGlyProAsnLeuThrSerSerPheHisValIleg1 288
Db      649 GGTGGCTTCTGATC-----ACCTCCAGC----- 673
Qy      288 yGluIlePheAspLysValHisPheGluGlyLysGlyValAsnHisValIleg1Lth 308
Db      674 -----GACGTGATCCAT-----TC 687
Qy      308 rThrLeuIleProAlaGlyAla-----AlaIleThrGluPhe----- 321
Db      688 CTGTGGTGTCCCGCTTCCGGGTCAAGCCGAGCCATCCCGGCTTCTCAACAGAGC 747
Qy      322 -----LysValAspValProGlyAspTyr----- 329
Db      748 CTGACCAAGGTCCAGACCCGCGCATCTATCGCGCAAGTGGCCGAGCTGCGGCA 807
Qy      330 -----ValLeuValAspHisAlaIlePheArgAlaPheAsnLysG1 343
Db      808 GACACGAGCTTCATCCGATCGTGTGAC----- 838
Qy      343 yAlaLeuGlyIleLeuLysValGlyGlyGluAsnHisGluIleTyrSerHisLysG1 363
Db      839 -----GTCAAGCCCAAGCGAGTTCACAGTGGCTGGCCAGGCAAGGA 885
Qy      363 nThrAspAlaValTyrLeuProGluGlyAlaProGluAlaIleAspThrGlnGluAlaPyr 383
Db      886 A---GAGCGCGCAAGGTCAAGAA-----CTGACCAAGCAAGAGTGGAC 927

```



QY 383 oLythrProAlaProAlaAsnLeuGlnGlnIleuValGlyLysAlaThrTyrAs 403  
DB 928 CAAG-----GAGAGTTGGTGGCGCGGAGCAAGAGTTACCA 966  
QY 403 pSerAsnCySAAlaAlaCyHisGlnProAspGlyLysGlyValProAsnAlaPheProR 423  
DB 967 CACCATCTGCGCGCTGCGACAGGCCAGAGCCGAGGCGATGCCCATGTTCCCGCG 1026  
QY 423 oLeuAlaAsnSerAspTyrLeuAsnAlaAspHisAlaArgAlaAlaSerIleValAlaAs 443  
DB 1027 GCTGAAGGTTCCAAATGCTCACCGGGCCCAAGGACACACTTGGAAGTGTCTTCA 1086  
QY 443 nGlyLeuSerGlyLysLysIleThrValAsnGlyAsnGlyTyrGlySerValMetProAl 462  
DB 1087 CGCGCGTCCCGGACCGGCATGGCGGCTTCGCGAAGCAG----- 1126  
QY 462 AlLeAlaLeuSerAspGlnGlnIleAlaAsnValIleThrTyrThrLeuAsnSerPheG 482  
DB 1127 -----CTCAACGAGGTGCGACCTGCGCGGTGATCCTACGAGCGCAACGCTGGG 1179  
QY 482 ValnLysGlyGlyGlnLeuSerAlaAspAspValAlaLysAlaLysThrLys 500  
DB 1180 CAACGACATGGCGACATGCTACCCCGAAAGACGTGCTGCTTACAGACAGAA 1234  
RESULT 16  
US-09-252-991A-646/c  
; Sequence 646, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 646  
; LENGTH: 1641  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-646  
Alignment Scores:  
Pred. No.: 2,85e-06 Length: 1641  
Score: 151.50 Matches: 87  
Percent Similarity: 31.83% Conservative: 40  
Best Local Similarity: 21.80% Mismatches: 110  
Query Match: 5.77% Indels: 162  
DB: 4 Gaps: 18  
US-10-088-045-2 (1-502) x US-09-252-991A-646 (1-1641)  
QY 132 HisProAspSerLysMetProHisAsnValAspPheHisAlaAlaThrGlyProGlyGly 151  
DB 1174 CACCTGTAT-----CCACATCTACACACTTCCGAGCCGAGACTGACCTGCAGGT 1124  
QY 152 -----GlyAlaGlnAlaSerPheThrAlaProGlyHisThrSerThrPheSerPhe 168  
DB 1123 CACCGGCTACCACTGGAAGTGCAGTACAGTACCTGG----- 1085  
QY 169 LysAlaLeuGlnProGlyLeuTyrValTyrHisCySAAlaValAlaProValGlyMetHis 188  
DB 1084 -----CCAGGACGTGAGTACTTCAGCAACCTGGGACACCCCGACAG 1043  
QY 189 IleAlaAsnGlyMetTyrGlyLeuIleLeuValGlnProLysGlyLeuProLysVal 208  
DB 1042 -----CCAGATCCACAA-----CCGCGCAGGC 1022

QY 208 LAspLysGlyTyrTyrValMetGlnGlyAspPheTyrThrLysGlyLysGlyGly 228  
DB 1021 GAAGCAGCAGCATTACTCTCGAGGTGAC----- 991  
QY 228 nGlyLeuGlnProPheAspMetGlyLysAlaIleArgGluAspAlaGlyTyrValAlaPh 248  
DB 991 ----- 991  
QY 248 eAsnGlySerValGlyAlaIleuThrGlyGluAsnAlaLeuLysAlaLysValGlyGly 268  
DB 990 -----GAGCGCTGTGTCTGCTCCGCTGGCAGCA 962  
QY 268 rValArgLeuPheValGlyAsnGlyGlyProAsnLeuThrSerSerPheHisValIleG 288  
DB 961 GGTGCGCTTCTGATC-----ACCTCAGC----- 937  
QY 288 yGluIlePheAspLysValHisPheGlnGlyGlyLysGlyGluAsnHisAsnIleGlnTh 308  
DB 936 -----GACGTGATTCAT-----TC 923  
QY 308 rThrLeuIleProAlaGlyGlyAla-----AlaIleThrGluPhe----- 321  
DB 922 CTGTGGGTGCTGCGGCTTCCGCTGCAAGCGGCACCGCATCCCGGCTTGTCAACAGGC 863  
QY 322 -----LysValAspValProGlyAspTyr----- 329  
DB 862 CTGACCAAGGTGCGACGAGCCGCGCATCTATGCGCGCCAGTGGCGGAGCTGCGGCA 803  
QY 330 -----ValLeuValAspHisAlaIlePheArgAlaPheAsnLysG 343  
DB 802 GGACCAAGGCTTCATGCCGATGCTGTCAC----- 772  
QY 343 yAlaLeuGlyIleLeuLysValGlnGlyGluAsnHisGlyIleTyrSerHisLysG 363  
DB 771 -----GTCAGCCCAAGCCGAGTTCACAGTGGCTGGCCAGCGCAGGA 725  
QY 363 nThrAspAlaValTyrLeuProGlyGlyAlaProGlnAlaIleAspThrGlnGlyAlaP 383  
DB 724 A-----GAGCGCGGAGGATCAAGAA-----CTGACCGACAGAGATGAGC 683  
QY 383 oLythrProAlaProAlaAsnLeuGlnGlnIleLysAlaGlyLysAlaThrTyrAs 403  
DB 682 CAAG-----GAAAGTGTGTTGGCGCGCGGCGCAAGAGTTACCA 644  
QY 403 pSerAsnCySAAlaAlaCyHisGlnProAspGlyLysGlyValProAsnAlaPheProR 423  
DB 643 CACCATCTGCGCGCTGCTGACAGGCCAGAGCCGAGGATGCGCGCATGTTCCCGCG 584  
QY 423 oLeuAlaAsnSerAspTyrLeuAsnAlaAspHisAlaArgAlaAlaSerIleValAlaAs 443  
DB 583 GCTGAAGGTTGGAAGATCTGACCGGGCCCAAGGACACACTTGGAAGTGTCTTCA 524  
QY 443 nGlyLeuSerGlyLysLysIleThrValAsnGlyAsnGlyTyrGlySerValMetProAl 462  
DB 523 CGCGTGGCGGACCGGCATGGCGGCTTCGCGAAGCAG----- 484  
QY 462 AlLeAlaLeuSerAspGlnGlnIleAlaAsnValIleThrTyrThrLeuAsnSerPheG 482  
DB 483 -----CTCAACGAGGTGCGACCTGCGCGGTGATCCTACGAGCGCAACGCTGGG 431  
QY 482 ValnLysGlyGlyGlnLeuSerAlaAspAspValAlaLysAlaLysThrLys 500  
DB 430 CAACGACATGGCGACATGCTACCCCGAAAGACGTGCTGCTTACAGACAGAA 376  
RESULT 17  
US-09-252-991A-4893  
; Sequence 4893, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136

```

; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 4893
; LENGTH: 1554
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-4893

```

## Alignment Scores:

```

Pred. No.: 3,77e-06 Length: 1554
Score: 150.00 Matches: 44
Percent Similarity: 48.61% Conservative: 26
Best Local Similarity: 30.56% Mismatches: 61
Query Match: 5.71% Indels: 13
DB: 4 Gaps: 5

```

US-10-088-045-2 (1-502) x US-09-252-991A-4893 (1-1554)

```

QY 361 HisLysGlnThrAspAlaValTyrLeuProGlu-GlyAlaProGlnAlaIleAspThrG1 380
    |||:|||||
DB 835 CATGCCAG-----CTACTGAAGTCCCTGCGCGCGCAAGACGACCT 879
QY 380 ngLualPro-----LysThrProAlaProAlaAsnLeuGlnGluGlnIleLysAlaG1 398
    |||:|||||
DB 880 GCCCATGCCGACGACGACGACGCGCTGCGACGACCTGACGCTGCGCGG 939
QY 398 yLysAlaThrTyrAspSerAsnCysAlaAlaCysHisGlnProAspGlyLysGlyAlaPr 418
    |||:|||||
DB 940 CGGCTCGGCTACCGCGAGTTCCTCTCCGACTGCCACCGCAAGACGCGCGCTCC 999
QY 418 oAsnAlaPheProProLeuAlaAsnSerAspTyrLeuAsnAlaAspHisAlaArgAla1 438
    |||:|||||
DB 1000 GGGCATGTTCCCGCGCTGCGCGCGCAACCCGACGCTGCTTCG-----GCCAACCGGAG 1053
QY 438 aSerIleValAlaAsnGlyLeuSerGlyLysIleThrValAsnGlyAsnGlnTyrGluSe 458
    |||:|||||
DB 1054 CAGCTACTGCATATACCTGACCGGCTGGAACCGCGGAAACCGCAACCGCATCGCG 1113
QY 458 rVal-----MetProAlaIleAla--LeuSerAspGlnGlnIleAlaAsnValIleTh 475
    |||:|||||
DB 1114 GGTCTACACCATGCGCGCTTCGCCGACGCTGAGAGACCGGAAATGCCAGATCCTCAG 1173
QY 475 rTyrThrLeuAsnSerPheGlyAsnLysGlyGlnLeuSerAlaAspAspValAlaLal 495
    |||:|||||
DB 1174 CTTGCTCCGACGAGTGGGGCAACGAGGTTCTGATCGATGATCCGCGCAGGTGAAGA 1233
QY 495 sAlaLysLys 498
    |||:|||||
DB 1234 ACTGCCGCGAG 1243

```

## RESULT 18

```

; US-09-252-991A-4910
; Sequence 4910, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 4910
; LENGTH: 2139

```

```

; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-4910

```

## Alignment Scores:

```

Pred. No.: 6.22e-06 Length: 2139
Score: 150.00 Matches: 44
Percent Similarity: 48.61% Conservative: 26
Best Local Similarity: 30.56% Mismatches: 61
Query Match: 5.71% Indels: 13
DB: 4 Gaps: 5

```

US-10-088-045-2 (1-502) x US-09-252-991A-4910 (1-2139)

```

QY 361 HisLysGlnThrAspAlaValTyrLeuProGlu-GlyAlaProGlnAlaIleAspThrG1 380
    |||:|||||
DB 942 CATGCCAG-----CTACTGAAGTCCCTGCGCGCGCAAGACGACCT 986
QY 380 ngLualPro-----LysThrProAlaProAlaAsnLeuGlnGluGlnIleLysAlaG1 398
    |||:|||||
DB 987 GCCCATGCCGACGACGACGCGCTGCGACGCGCTGACCTGACGCTGCGGGG 1046
QY 398 yLysAlaThrTyrAspSerAsnCysAlaAlaCysHisGlnProAspGlyLysGlyAlaPr 418
    |||:|||||
DB 1047 CGGCTCGGCTACCGCGAGTTCCTCTCCGACTGCCACCGCAAGACGCGCGCTCC 1106
QY 418 oAsnAlaPheProProLeuAlaAsnSerAspTyrLeuAsnAlaAspHisAlaArgAla1 438
    |||:|||||
DB 1107 GGGCATGTTCCCGCGCTGCGCGCGCAACCCGACGCTGCTTCG-----GCCAACCGGAG 1160
QY 438 aSerIleValAlaAsnGlyLeuSerGlyLysIleThrValAsnGlyAsnGlnTyrGluSe 458
    |||:|||||
DB 1161 CAGCTACTGCATATACCTGACCGGCTGGAACCGCGGAAACCGCAACCGCATCGCG 1220
QY 458 rVal-----MetProAlaIleAla--LeuSerAspGlnGlnIleAlaAsnValIleTh 475
    |||:|||||
DB 1221 GGTCTACACCATGCGCGCTTCGCCGACGCTGAGAGACCGGAAATGCCAGATCCTCAG 1280
QY 475 rTyrThrLeuAsnSerPheGlyAsnLysGlyGlnLeuSerAlaAspAspValAlaLal 495
    |||:|||||
DB 1281 CTTGCTCCGACGAGTGGGGCAACGAGGTTCTGATCGATGATCCGCGCAGGTGAAGA 1340
QY 495 sAlaLysLys 498
    |||:|||||
DB 1341 ACTGCCGCGAG 1350

```

## RESULT 19

```

; US-09-252-991A-8339
; Sequence 8339, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 8339
; LENGTH: 1977
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-8339

```

## Alignment Scores:

```

Pred. No.: 1.01e-05 Length: 1977
Score: 147.50 Matches: 85
Percent Similarity: 34.99% Conservative: 56
Best Local Similarity: 21.09% Mismatches: 147

```

Query Match:	5.62%	Indels:	115
DB:	4	Gaps:	20

Oy	107	ThrPheGlyGlyValProGlyIleMetIleArgValArgGlyGlyAspThrIleGlu	126
Db	328	ACCATCAACGGTTCCGCTGCGGCGCCGACCCCTGGCGGGAAGGCGACAACGTGAC	387
Oy	127	ValGlnPheSerAsnHis-----ProAspSerLys-----Met	137
Db	388	CTGGCGGGTGGCGAACCGGCTCGCGGACGACACTCGATGCCATGGCAGGATATCTGG	447
Oy	138	ProHisAsnValAspPheHisAlaIleThrGlyProGlyGlyValAlaGluIleAspPhe	157
Db	448	CCGGCGCAACATGGAT-----GCGGTGCGGGCGCTTCAGCTTCGAGGGC-----	489
Oy	158	ThrIleAspGlyHisIleThrSerThrPheSerPheLysAlaLeuGlnProGlyLeuThrVal	177
Db	490	ATCGCTCTTGGGGCTCTTCAAGATACCGCTTCAGAGTCCGGCAACGGCAACCTTAGCG	549
Oy	178	TyrHisCysAlaValAlaProValGlyMetHisIleAlaAsnGlyMetTyrGlyLeuIle	197
Db	550	TACCAAC-----AGCCACTCGGGGCTACAGAGACAGCGCGGGGTCTACGGCGCCCTG	600
Oy	198	LeuValGluProLysGlyGlyLeuPro--LysValAspLysGlyTyrTyrValMetGln	216
Db	601	GTCATTCAGCGCCCGGAGCCGGAACCTTCACTCACTGCAGCCGGCACTACGTGTGCTGCTC	660
Oy	217	GlyAspPheTyrThrLysGlyLysTyrGlyGlyGlnGlyLeuGlnProPheAspMetGlu	236
Db	661	AGCAGCTGCTGGACGAC-----AAGCCGACGGCATCTCTC	696
Oy	237	LysAlaIleArgGlyAspAlaGlyTyrValValPhe-----	248
Db	697	GCCAAAGCTGAAGAGCACTCCGACTATCAACTTCCACAAGCGACACGTCCGCGACTTC	756
Oy	249	-----AsnGlySerValGlyAlaLeuThrGlyGluAsn-----	259
Db	757	ATCGACGACGTGACGGCCAGCGGTGGGCGCGGACCTTGGCCGACCGAAGATGTGGCC	816
Oy	260	AlaLeuLys-----AlaLysValGlyGlyThrValArgLeuPheVal	273
Db	817	GAGATGAAGATGAGTCCGACCGATCTCGCCAGCGTCAAGCGGTACACTATATCACTG	876
Oy	274	GlyAsnGlyGlyPro-----AsnLeuThrSerSerPheHisValIleGlyGluIle	290
Db	877	CTCAACGGACGACCGCGCGACGCGAATGACGCGCGCTGTTCCGC-----	921
Oy	291	PheAspLysValHisPheGluGlyGlyLysGlyGlyAsnHisAsnIleGlnThrThrLeu	310
Db	922	-----CCGGCGGAAGAGCTCCGCTCGCTTC-----	948
Oy	311	IleProAlaGlyGlyAlaAlaIleThrGluPheLysValAspValProGly-----	327
Db	949	-----GTCAACGCGCTCGCGCATGAGCTATTTCCAGCGTCCGATTCGCCGCGCTGAAGATG	1002
Oy	328	-----AspTyrValLeuValAspHis-----	334
Db	1003	ACCGTGTGGCGCGGACGACGACGACGTCGACCGCGCTCAGCGTGAAGATGTGGCATC	1062
Oy	335	AlaIlePheArgAlaPheAsnLysGlyAlaLeuGlyIleLeuLysValGluGlyGluGlu	354
Db	1063	GCGGCGCGGACCTTACGACGTG-----ATCGTGAACCGGCGCGGACGCGC	1110
Oy	355	AsnHisGluIleTyrSerHisLysGlnThrAspAlaValTyr-----	368
Db	1111	GCGTATACCTCTTGGCGCGAGTGCATGACCGCGACGGGTACGCGCGGACACCTGGGG	1176
Oy	369	LeuProGluGlyAlaProGlnAlaIleAspThrGlnGlyAlaProLysThrProAlaPro	388
Db	1171	CTGGCGGAAGGCTTAGCGCGCGGTG-----CCAGCGCCGATCTCG	1212

[illegible]

```

RESULT 20
US-09-252-991A-14102
; Sequence 14102, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 14102
; LENGTH: 1782
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-14102

```

Alignment Scores:	
Pred. No.:	1,24e+05
Score:	146.00
Percent Similarity:	33.78%
Best Local Similarity:	22.1%
Query Match:	5.56%
DB:	.4
	Gaps: 25
	Indels: 128
	Mismatches: 219
	Conservative: 61
	Matches: 116
	Length: 1782

US-10-098-045-2 (1-502) x US-09-252-991A-14102 (1-1782)

Oy 68 HisAlaProGluValProProProValAspArgSerHisProAlaLysValValLys 87  
          |||         |||         |||         |||         |||         |||  
Db 274 CATTCCCGAGCTGAACCCCATCACAGAAGTCTTCGACAGATGCCCTTCCGTGCAG 333  
          ||||         |||         |||         |||         |||         |||  
Oy 88 MetGluThrValGluLysValMetArgIleuAlaSerGlyValGluTyrglnPheThr 107  
          |||         |||         |||         |||         |||         |||  
Db 334 CTG-----TTGGCGCGCGGGCTTGCTCAAGCGCCACCGT 368  
          |||         |||         |||         |||         |||         |||  
Oy 108 PheGlyGlnValProGly-----GlnMetLeuArgValArgGluGlyAspThr 124  
          |||         |||         |||         |||         |||         |||  
Db 370 GCTGGTGGAAAAAGCCACGCCGACCAGCATCAAGCGCAGCAGCTGGACGAGCGATCGA 428  
          |||         |||         |||         |||         |||         |||  
Oy 125 IleGluValGlnPheSerAsnHisProAspSerLysMetProHisAsnValAspPheHis 144  
          |||         |||         |||         |||         |||         |||  
Db 430 GGACACCCCTGGCGCATCATCTGCGCTTGACACGCGACTACGCGTACTACAACGCCAC 488  
          |||         |||         |||         |||         |||         |||  
Oy 145 -----AlaIleThrGlyProGly----GlyGlyAlaGluAlaSerPheThrAla 155  
          |||         |||         |||         |||         |||         |||  
Db 490 CGCGCAGTGTCTGACGAATCTGGCTGCTCATGAAGAGGCTTGAGAGATGAAGCGGACCTT 549  
          |||         |||         |||         |||         |||         |||  
Oy 160 ProGlyHisIleThrSerThrPheSerPheLysAlaLeuGlnPro----- 173  
          |||         |||         |||         |||         |||         |||  
Db 550 TCCGGC-----CTGCCCTGGCGCTGGCGCGCGCTGTGCCGCTTCGAGCTCCGCCGCC 603  
          |||         |||         |||         |||         |||         |||  
Oy 174 -----GlyLeuTyrValTyrlHis----- 179

```

Db      604 GCGCCGACGCGCGCTGTCAGCGCGGCGAATATCTCCGCCGCCGCGCATGCTGATG 663
Qy      180 ---CysAlaValAlaProValGlyMetHisIleAlaAsnGlyMet----- 193
Db      664 GCTTCCCATACCGCGAGGCGCGCGCGCTTCCGCCGCCGCGCTGCGCATGCTGCGCCG 723
Qy      194 -----TyrGlyLeuIleValGluProGlyGlu-----GlyLeuPro 206
Db      724 TTGGGACCATCTACGGGACCAACATCAACCCGACACAGAAACAGGATCGGCGCTAC 783
Qy      207 LysValAspLysGlyTyrTyrValMetGlnGlyAspPheTyrThrLysGlyLysTyrGly 226
Db      784 AGCCGCCACGAGTCTTCCGCGCTCTT-----ACCGAGGCGCAACGCGAG 828
Qy      227 GluGlnGly-----LeuGlnProPheAspMetGluLysAlaIle---ArgGlu 241
Db      829 GACGCGCCCTACCTCTATCCGCGCATGCCCTTACCTCTATCACTGATCGAGCGCGAG 888
Qy      242 AspAlaGluTyrValVal-----PheAsnGlySerValGlyAlaLeuThrGlyLysAla 260
Db      889 GATCCGACCGCATCTACGCTTACCTGATGCGCCACGAGCCGATCGCCGCCGCCGCG 948
Qy      248 -----PheAsnGlySerValGlyAlaLeuThrGlyLysAla 260
Db      949 CAGACGACCTGACCTTCCGCTTACGTCGCGCATGGGC---CTGGCCGCGCTGGAACCTG 1005
Qy      261 LeuLysAlaLysValGlyGluThrValArgLeuPheValGlyAsnGlyGlyProAsnLeu 280
Db      1006 CTC-----TACGCGAAGAGGCTGGCTTGCAGCGCGAGGAGGAGCGCAGGAGCC 1056
Qy      281 ThreSerPheHisValIleGlyGluIlePheAspLysValHisPheGluGlyLys 300
Db      1057 TGGAAAGCGCGGACGATACGTGTGAGGTGCTC-----GGCAGCTGC 1098
Qy      301 GlyGluAsnHisValIleGlnThrLeuIleProAlaGlyAlaAlaIleThrGlu 320
Db      1099 GCGGAGTGCATATCCCGCGCAC-----CTGGCGCGCGCGCTGGAGACAGAC 1146
Qy      321 PheLysValAsp-----ValProGlyAsp 328
Db      1147 AAGCGCTGAGCGCGCGCTGCTCAACGCGCTACCTGGCGCGCGCTGCTGCGCGAGC 1206
Qy      329 -----TyrValLeuValAspHisAlaIlePheAsnGlyAla 343
Db      1207 CTGGCGCGCGCGCTTGGACCCAGCGGACCTGCGGAGCTTCTCAAGACAGCATAGC 1266
Qy      344 AlaLeuGlyIleLeu-----LysValGluGlyGluAsnHisGlu 357
Db      1267 GCCCAGGCGAGCATGTTCAACGAGATGTTCCGGTGTGACACACAGCAGCAGCAGCTC 1326
Qy      358 IleTyrSerHisLysGlnThrAspAlaValTyr---LeuProGluGlyAlaProGlnAla 376
Db      1327 GACGACAGCAGCCTGGACCGCATGCGACCTGCTCGGCGACGCGCGCGCGCGC 1386
Qy      377 IleAspThrGlnGluAlaProLysThrProAlaProAlaAsnLeuGlnGlnIleLys 396
Db      1387 AAGTGTCTCAGCGCGCTGCCGAG-----GCCGAGCTGAACGACAGCGGAG 1434
Qy      397 AlaGlyLysAlaThrTyrAspSerAsnCysAlaAlaCysHisGlnProAspGlyLysGly 416
Db      1435 CGCGGTGCGCAGCATCTCAACGTCGTGCGCGCGCTGCCACGCGCGCGGAGGAGGCG 1494
Qy      417 ValProAsn---AlaPheProLeuAlaAsnSerAspTyrLeuAsnAlaAspHisAla 435
Db      1495 AAGCGCATATCGCGGTGCGCATGACGCGCAACCTACCTGCGCGCTGCAAGACCGCGT 1554
Qy      436 ArgAlaIaSerIleValAlaAsnGlyLeuSerGlyLysIleThrValAsnGlyAsnGln 455
Db      1555 AATGTGTCGAGATGAGAGGAGCATCTCAACAG---CAGTTCACCGCGCTTGAA 1611
Qy      456 TyrGluSerValMetProAlaIleAla---LeuSerAspGlnGlnIleAlaAsnVal 473

```

```

Db      1612 CGCATGCAACCGATGCCCGGTTTCGCCGGGAACTGACAGCAGCACTACGAGCTG 1671
Qy      474 IleThrTyrThrLeuAsnSerPheGlyAsnLysGlyGlnLeuSerAlaAspAspVal 493
Db      1672 CTCACTACTCTGGCGGACAGACTGGGCGGATTCGCCGGTACTCGGTCCCAACAGGTC 1731
Qy      494 AlaLysAlaLys 497
Db      1732 GCGCAATTGAG 1743

```

## RESULT 21

US-08-706-037-24  
Sequence 24, Application US/08706037  
Patent No. 5770419

## GENERAL INFORMATION:

APPLICANT: Xu, Feng  
APPLICANT: Betka, Randy M.  
APPLICANT: Wahleithner, Jill A.  
TITLE OF INVENTION: BLUE COPPER OXIDASE MUTANTS WITH  
TITLE OF INVENTION: ENHANCED ACTIVITY  
NUMBER OF SEQUENCES: 29  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 57704190 No. 5770419disk of No. 5770419th America, Inc.  
STREET: 405 Lexington Avenue, 64th Floor  
CITY: New York  
STATE: New York  
COUNTRY: United States of America  
ZIP: 10174-6401

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/706,037  
FILING DATE: 30-AUG-1996  
CLASSIFICATION: 435

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/002,800  
FILING DATE: 1-SEP-1996

ATTORNEY/AGENT INFORMATION:  
NAME: Harrington, James J.

REGISTRATION NUMBER: 38,711  
REFERENCE/DOCKET NUMBER: 4526.200-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-867-0123  
TELEFAX: 212-878-9655

INFORMATION FOR SEQ ID NO: 24:

SEQUENCE CHARACTERISTICS:

LENGTH: 1588 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

FEATURE:

NAME/KEY: CDS

LOCATION: 1..1587  
US-08-706-037-24

## Alignment Scores:

Pred. No.:	3,49e-05	Length:	1588
Score:	141.00	Matches:	103
Percent Similarity:	32.28%	Conservative:	60
Best Local Similarity:	20.40%	Mismatches:	216
Query Match:	5.37%	Indels:	126
DB:	1	Gaps:	20

US-10-088-045-2 (1-502) x US-08-706-037-24 (1-1588)

```

Qy      59 LeuProValIleAspAlaIleValThrHisAlaProGluValProProProValaAspArg 78
Db      22 CTACCTTCTCTGCTGCGGTCTCAAC----- 48

```

QY 79 AsphHisProAlaIaIyValIaIyValIyS-----MetGluThrValGluIuIyValMetArg 96  
DB 49 -----CCGCGCTTCCTGCGCGCGCACTATATAGTTGCGCACTCAAGCAAGCGCATATGTC 102  
QY 97 LeuAlaAspGluValGluIuIyrglnPheThrPheGluGluIuIyValProGluIuIyMet 116  
DB 103 GCTCCGGAATGGCTTTTACAGCGCTTATGCTTCGTCGAACGGTTTATGTTCTCTGCGACGTTG 162  
QY 117 IleArgValArgGluIuIyAspThrIleGluValGlnPheSerAsnHis-----Pro 133  
DB 163 ATCAGCGCGCAACAGGAGTGACACCTTGCGCATTAATGTCAGATCAAGTCAACGCGACCT 222  
QY 134 AspSerLysMetProHisAsnValAspPheHis-----AlaAlaThrGluPro 149  
DB 223 AGTATGCGCGTGGCCCAACAGATTCATGTCATGATGTCATGTCATCACTACTACCGCGCAC 282  
QY 150 GluIuIyGluValGluIuIaIaIaSerPheThrAlaPro-----GlyHisThrSer 164  
DB 283 GAGGATGGCGCGCATTCCTCAACGCAATGCCCTATGTCGCAAAATTTGCTCATATCATAC 342  
QY 165 ThrPheSerPheLysAlaLeuGlnProGluIuIyLeuIyValIyThrHisCysAlaValAlaPro 184  
DB 343 GAGATCCCATTTGGCGGCGC---CAACAGGAGACCATGTCGATAC-----GCCCAT 390  
QY 185 ValGluMetHisIleAlaAsnGluMetIyrgly---LeuIleLeuValGluPro----- 201  
DB 391 CTTCGAGTCATATATGTCATGATGATTCGAGCGCTTGTGTCATCATATGATCCAAACGAC 450  
QY 202 -----LysGluGluIuIyLeuProLysValAspLysGluIyValMetGlnGluAspPhe 219  
DB 451 CCACACAGATCGCGCTACAGACGTCGATGATGCGAGACAGCATAGTCATGCTTAGAGACTGG 510  
QY 220 Tyr-----ThrLysGluIyLysIyrglyGluIuIy----- 229  
DB 511 TACCATCTCCGCGACCCCTTCTAGAAAAGCAATTTCTCCAGCTATATACACCGCTCG 570  
QY 230 LeuGlnProPheAspMetGluLysAlaIleArgGluAspAlaGluIyValIuIyPheAsn 249  
DB 571 CTCTCTCCCTGTTCCGGACTCGGGTCTTATCAATGGCAAGCGCGCTATGTG----- 621  
QY 250 GlySerValGluAlaLeuThrGluIuIaIaIaLeuLysAlaIyValGluIuIyThrVal 269  
DB 622 -----GGCGTCCCGCGACGTCCTCCCGTCAGTATACGTAACGTAAACGTGGAACCATAT 675  
QY 270 ArgLeuPheValGluIyAsnGluIyGluProAsnLeuThrSerSerPheHisValIleGluIu 289  
DB 676 CGCTTCGCGGTATATCAACGCTTCTGTCATCGGTCGCTTACCTTTTCATC----- 726  
QY 290 IlePheAspLysValHisPheGluGluIyGluIyGluIuAsnHisAsnIleGlnThrThr 309  
DB 727 -----GAGGACATATGTCG-----ACT 744  
QY 310 LeuIleProAlaGluIy-----AlaAlaIleThrGluPheLysValAsp 324  
DB 745 GTCATTGAGCGCGATGCGATCTCTGACACAGCCCTTGCGTGTGACAGCTTCAGATTTAC 804  
QY 325 ValProGluIyAspIyValLeuValAspHisAlaIlePheArgAlaPheAsnLysGluIa 344  
DB 805 GGTGACACAGCTACTCTCTCTCATCTGTGAAGCCACCAACGCGCGCACTACTGCGATT 864  
QY 345 LeuGluIleLeuLysValGluIuIyGluIuIaIaIaGluIleIySerHisLysGlnThr 364  
DB 865 CGTGACACCATGACCGTTGACGAGCGGAGCCGAACATGCAAACTTGAGCCCAACCATATGTC 924  
QY 365 AspAlaValIyLeuProGluIuIyAlaProGlnAlaIleAspThrGlnIu----- 381  
DB 925 TTTCGCTGATTTGCACTACGAGGAGCGCCCAACGCCGAACCCACGAGCAACAGGCACT 984  
QY 382 -----AlaPro 383  
DB 985 GGTATCGGTACTGCACTGTTGAAGGAACCTCCATGCGCTATCAACCTTGCGCTCG 1044  
QY 384 LysThrProAlaProAlaAsnLeuGlnGluIuIleLysAlaGluIySAlaThrIyAsp 403

DB 1045 GCGGCGCTCCGCTCCGCGACAGCTTTCCTCAATCTTTCGATTTGGCAATTTGGCGCGCACAGTTGAT 1104  
QY 404 Ser-----AsnCysAlaIaIaCysHisGlnProAspGluIySglVal 417  
DB 1105 GGGATTTTAGGTTACATTTATATACATCAAGTACAGAGCTCT----- 1149  
QY 418 ProAsnAlaPheProProLeuAlaAsnSerAspIyLysAlaAsnAlaAspHisAlaArgAla 437  
DB 1150 -----TCGTCGCCACGCTC----- 1164  
QY 438 AlaSerIleValAlaAsnGluIuIySerGluIyIleThrValAsnGluIyAsnGlnIyrglu 457  
DB 1165 TTGAAGATTTTGGCAACATGCGAGCATGACGCCGATTTTCACGCCCAATGACGACACT 1224  
QY 458 SerValMetProAlaIleAlaLeuSerAspGlnIleAlaAsnValIleThrIyThr 477  
DB 1225 ATCGTATTTGCCACACATTAAGTTATCGAGCTCAATATATACCGGAGGTGCAGACACCT 1284  
QY 478 LeuAsnSerPheGluIyAsnLysGluIuIyGlnLeuSerAlaAspAspValAlaLysAlaLys 497  
DB 1285 ATCCATCTCCACGCGCAT-----GTGTTGATATGTCGTAATTCACAC 1326  
QY 498 LysThrLysProAsn 502  
DB 1327 GGTGCTACCCCGAAC 1341  
RESULT 22  
US-09-005-397-24  
Sequence 24, Application US/09005397  
Patent No. 5972670  
GENERAL INFORMATION:  
APPLICANT: Xu, Feng  
APPLICANT: Berka, Randy M.  
TITLE OF INVENTION: BLUE COPPER OXIDASE MUTANTS WITH  
TITLE OF INVENTION: ENHANCED ACTIVITY  
NUMBER OF SEQUENCES: 29  
CORRESPONDENCE ADDRESS:  
ADDRESSER: No. 59726700 No. 5972670disk of No. 5972670th America, Inc.  
STREET: 405 Lexington Avenue, 64th Floor  
CITY: New York  
STATE: New York  
COUNTRY: United States of America  
ZIP: 10174-6401  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/005,397  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/706,037  
FILING DATE: 30-AUG-1996  
APPLICATION NUMBER: US 60/002,800  
FILING DATE: 1-SEP-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Harrington, James J.  
REGISTRATION NUMBER: 38,711  
REFERENCE/DOCKET NUMBER: 4526.200-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-867-0123  
TELEFAX: 212-878-9655  
INFORMATION FOR SEQ ID NO: 24:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1588 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)





```

Sequence 208, Application US/09615192A
Patent No. 6410718
GENERAL INFORMATION:
APPLICANT: Bloksberg, Leonard N.
APPLICANT: Havukkala, Ilkka
TITLE OF INVENTION: Materials and Methods for the
FILE REFERENCE: 11000.1003c4U
CURRENT APPLICATION NUMBER: US/09/615,192A
CURRENT FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 08/975,316
PRIOR FILING DATE: 1997-11-21
PRIOR APPLICATION NUMBER: US 08/713,000
PRIOR FILING DATE: 1996-09-11
PRIOR APPLICATION NUMBER: US 09/169,789
PRIOR FILING DATE: 1998-10-09
NUMBER OF SEQ ID NOS: 405
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 208
LENGTH: 980
TYPE: DNA
ORGANISM: Eucalyptus grandis
US-09-615-192A-208

```

Alignment Scores:		
Pred. No.:	4,91e-05	Length:
Score:	136.71	Matches:
Percent Similarity:	34.51%	Conservative:
Best Local Similarity:	20.94%	Mismatches:
Query Match:	5.20%	Indels:
DB:	4	Gaps:

```

QY 78 AAGSAPRIVRQIALEVALVALVALVALVALVALVALVALVALVALVALVALVALVALMEKCYDEU 97
Db 27 CGTAACTACACATTCATTAATGTGTAAAGAAATACACG-----AGACTG 71
QY 98 AIAAPBGLVALGLUTYTGlnPheThrPheGlnGlnValProGlnMeIle 113
Db 72 TGTTCG-----AGCAAGCCCATCGTAAACGGGATGTTCCCGGACCCCATCTC 122
QY 118 ARGVALARGGLUGLVARThrTlleglVALGlnPheSerAsnHisPProAspSerIleMet 133
Db 126 TATGCTAAGGAAATATACACACCGTGTCTGAGGGTCTCTAAC-----CGTCTC 172
QY 138 ProHisAsnValAspPheHis-----AlaIleThr 141
Db 174 AAATACATATCTACCATCCATTCGGCATGTGATCCGGCATTCAGACAGCGGGTGGGCGCAC 233
QY 148 GLYProGlnGLYGLVALAGLUALAspPheThrAlaProGlnHisThrSerPheSer 167
Db 234 GGGCCACGACATACATTTCACCATGCGCGGATC-----CAGCGGGGCCAAAGCATGTGTACAT 290
QY 168 PheIleValIleu---GlnProGlnLeuTYGValTYGHisCysAlaValAlaProValGly 186
Db 291 TTTCACCATCACGGGGCCACAGGGGCAACCTCTGTGGCAATGCACATATA----- 330
QY 187 MetHisIleIleAlaAsnGlyMetTYGValIleValIleValGluProGlnGlnGlyLeu--- 209
Db 339 CTGTGGCTCAGGGCAACCTTCGACGGAGCCATTTGATCTTATTCGCCCAAGCGTGTGTCTCA 399
QY 206 -----ProIleValIleAspIlySGlUTYTYGValMetGlnGlnIAspPheTYRThrIlys 222
Db 399 TACCCCTTCCTTAACCCCAACAAAGAAAGTTGTGTGTCTATTGGGGGAATGGCGAAATCT 455
QY 223 GLYLeuTYGGLY-----GlnGlnGlnGlyLeuGlnProPheAspMetGlu 233
Db 459 GATACAGAAAGGTGTGATCATGTCAAAGCATCAAGTCCGAGTATGACACCG----- 506
QY 237 LyeAlaIleArgGlnIleAspIleAGlUTYGVAlValPheAsnGlySerValGlyAlaLeuThr 255
Db 507 -----AATGCTTCGAGTGTCAACAGATCAAGTAAAGGCGATCCAGGGCGCAATGCC 555

```

[illegible]

```

RESULT 25
US-09-252-991A-8410
: Sequence 8410, Application US/09252991A
: Patent No. 6551795
: GENERAL INFORMATION:
: APPLICANT: Marc J. Rubenfield et al.
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
: TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
: FILE REFERENCE: 107196.136
: CURRENT APPLICATION NUMBER: US/09/252,991A
: PRIOR FILING DATE: 1999-02-18
: PRIOR APPLICATION NUMBER: US 60/074,788
: PRIOR FILING DATE: 1998-02-18
: PRIOR APPLICATION NUMBER: US 60/094,190
: PRIOR FILING DATE: 1998-07-27
: NUMBER OF SEQ ID NOS: 33142
: SEQ ID NO 8410
: LENGTH: 1203
: TYPE: DNA
: ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-8410

```

Alignment Scores:	
Pred. No.:	0.000179
Length:	1201
Score:	132.50
Matches:	45
Percent Similarity:	42.78%
Conservative:	32
Mismatches:	64
Best Local Similarity:	25.00%
Query Match:	5.05%
Indels:	39
DB:	4
Gaps:	8

US-10-088-043-2 (1-502) x US-09-252-991A-8410 (1-1203)

Qy	107	ThrPheGlyGlyValProGlyGlyMetLeaValAArgGlyAspThrIleGlu	126
		:::	:::
Db	344	ACCATCAACGGATTGGCTGCCGGCCGACCTGGCGTGGCGAAGGGCAACGTGACC	403
		:::	:::
Qy	127	ValGlnPheSerAsnHis-----ProAspSerLys-----Met	133
		:::	:::
Db	404	CTGCGGGTGGCAACCGCTCCGACGAGACACTGTGATCACTGCAACGGCATCATCTG	463
		:::	:::
Qy	138	ProHisAsnValAspPheHisIalaIaThrGlyProGlyGlyValAlaGluAlaSerPhe	157
		:::	:::
Db	464	CCGGCCAAATGGAT-----GACGTCCGGGGCTTCAGCTTCGAAAGGC-----	505





Db 1447 CAATCTGAGAAATAGCTTAAAGATGACCA 1479

RESULT 27  
US-09-328-352-3903  
Sequence 3903, Application US/09328352  
Patent No. 6562958  
GENERAL INFORMATION:  
APPLICANT: Gary L. Breton et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER.  
FILE REFERENCE: GTC99-03PA  
CURRENT APPLICATION NUMBER: US/09/328,352  
CURRENT FILING DATE: 1999-06-04  
NUMBER OF SEQ ID NOS: 8252  
SEQ ID NO 3903  
LENGTH: 1908  
TYPE: DNA  
ORGANISM: Acinetobacter baumannii  
US-09-328-352-3903

Alignment Scores:  
Pred. No.: 0.000471 Length: 1908  
Score: 131.50 Matches: 73  
Percent Similarity: 35.36% Conservative: 49  
Best Local Similarity: 21.16% Mismatches: 116  
Query Match: 5.01% Indels: 107  
Gaps: 19

US-10-088-045-2 (1-502) x US-09-328-352-3903 (1-1908)

Qy 107 ThrPheGlyGlyGlnValProGlyGlnMetIleArgValArgGlyGlyAspThrIleGlu 126  
Db 142 ACTGTAATGTAAGTTGTTGACCTCTTCTGATGTTGAGAGGATGATGATGCCGT 201  
Qy 127 ValGlnPheSerAen-----HisProAspSerIleMetProHisAsnValAspPhe 143  
Db 202 ATTCGCGTACACCAACAGTTGAAAAACCAAGATTCA-----TCTATTCACTGG 249  
Qy 144 HisAlaIaIaThrGlyProGly-----GlyGlyAlaGluAlaSerPhe 157  
Db 250 CATGCTTAATTTATTAACCTGATGATGATGATGATGATGATGATGATGATGATGAT 309  
Qy 158 ThrIaProGlyHisThrSerThrPheSerPheIleValLeuGlnProGlyLeuTyVal 177  
Db 310 ATTCACCAATTAACCTATGATATTAATTAATTAATTAATTAATTAATTAATTAAT 369  
Qy 178 TyrHisCysAlaValAlaProValGlyMetHisIle-----AlaAsnGlyMet 193  
Db 370 TATACAC-----TCGCATGTAAGCAAGCCCAAGCAAGCAAGATGCGCTTG 408  
Qy 194 TyrGlyLeuIleLeuValGluProGlyGlyLeuPro-----LysVal 208  
Db 409 TATGCTCTAGTTATTTATCCCAAAATAAAGTTCTTTAAGCGCTGAGAAAAAGCT 468  
Qy 209 AspIleGlyGlyTyValMetGlnGlyAspPheTy-----ThrIleGlyLys----- 224  
Db 469 GATAGAGATTAAGTATGATCTTCTGATTTTCAATCTGATGATGATGATGATGATGAT 528  
Qy 225 -----TyrGlyGlnGlnGlyLeuGlnProPheAspMet 235  
Db 529 AGTAATCTCAAGAAAGAGCTGATTTATTAACCAATCGAAGCAAGCAAGCTTATTAAGT 588  
Qy 236 GluIleValAlaIleArgGlyAsp-----ValPheAsn 249  
Db 589 TTCAAGCAGATTAAAGAGATGTTTAAAGCAACATGAAAAAGATCGTTCCATGTCGAT 648  
Qy 243 -----AlaGlyTyVal-----ValPheAsn 249  
Db 649 CAGATGCGAATGCTTAAGACAGATATGCTGATGTAACAACACTADAGCTTTTATGAT 708  
Qy 250 GlySerValGlyAlaLeuThrGlyGluAsnAlaLeuLysAlaLysValGlyGlnThrVal 269  
Db 709 GGTAAAGACTCCAGAAACAAATTTGAGCTGAAATTTTAAAGCT-----GGAGAAAAAGTA 762

Qy 270 ArgLeu-----PheValGlyAsnGlyGlyProAsnLeuThrSerSerPheHisValIleGly 288  
Db 763 GCTTCTGCTTTATTCATGATGTTGCGCAATGCTTTG----- 798  
Qy 289 GluIlePheAspLysValHisPheGlyGlyGlyGlyGluAsnHisAsnIleGlnThr 308  
Db 799 -----TTCATGTA-----AGAATTCCTAATCGAAATG 828  
Qy 309 ThrIleuIleProAlaGly-----AlaAlaIleThrGluPheLysVal 323  
Db 829 ACAATAGTACCGCGGATGACCAACAGTAAACCCGTGCAATGATGATGATGATGATGAT 888  
Qy 324 AspValProGlyAspTy-----ValLeuValAsp-----HisAlaIlePheArg----- 338  
Db 889 GGAACCGCTGACATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 948  
Qy 339 -----AlaPheAsnLysGlyAlaLeuGlyIleLeuLysValGluGlyGluAsnHis 356  
Db 949 GCTGAATCTATGATGCAAGTGGTTTCTATCGTACATTAATGATGATGATGATGATGAT 1008  
Qy 357 GluIleTySerHisLysGlnThrAspAlaValTyLeuProGlyGlyAlaProGlnAla 376  
Db 1009 CTTGTAATAAAT-----ATATGATGCCGACACCTCGTCCCGCTGCT 1050  
Qy 377 IleAspThrGlnGlu 381  
Db 1051 TTACTGACGATGGA 1065

RESULT 28  
US-09-252-991A-14169/c  
Sequence 14169, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 14169  
LENGTH: 1323  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-14169

Alignment Scores:  
Pred. No.: 0.000339 Length: 1323  
Score: 130.50 Matches: 82  
Percent Similarity: 35.51% Conservative: 70  
Best Local Similarity: 19.16% Mismatches: 152  
Query Match: 4.97% Indels: 124  
Gaps: 19

US-10-088-045-2 (1-502) x US-09-252-991A-14169 (1-1323)

Qy 117 IleArgValArgGlyGlyAspThrIleGluVal-----GlnPheSerAenHis 132  
Db 1108 GTTCGCCGCGGCGGCTCCGATTCAGTCCGCCGCTTCCGACACATCTAGCGCAACCAATC 1049  
Qy 133 -ProAspSerLysMetProHisAsnVal-----AspPheHisAlaIle 146  
Db 1048 CCCGACCAAGAA-----CACGCATCGCGCGCTACAGCGCGCAAGAGATTCTTCCGCC 995  
Qy 146 aThrGlyProGlyGlyAlaGluAlaSerPheThrAlaProGlyHisThrSerThrPhe 166  
Db 994 TCTTACGAGGCGCAAGCGGACGCGCTTACTCTATCCGCGCATGCCCTTACACC-- 937



DB 1443 GTGAACAGCCGCTCCATCCCGGCTTACGCTGATCCCGAAGATGACTACACTGTC 1502  
QY 215 MetGlnGlyAspPheThrThrGlyGlySerGlyGlnGlnGlyLeuGlnProPhe--- 233  
DB 1503 CTCATCGGTGACTGCTACCTAAG-----AGCCACACCCAGTTGAAAGAGTTCTTC 1553  
QY 234 AspMetGlnGlyValAlaIle---ArgGlnAspAlaGlnGlyValValPheAsnGlySerVal 252  
DB 1554 GACGCTGCTCTACTATTGCTGCTCCAGACGCT-----ATTGCTATCAACGGAAGTCC 1607  
QY 253 GAlaAlaLeuThrGlyGlnAsnAla-----LeuGlyAlaValGlyGlnThrValArg 270  
DB 1608 GGAAGAGGTGATGATCAGACGACCGCTTTCACCTTGAAGCTCGAAGAACTTACAGG 1667  
QY 271 LeuPheValGlyAsnGlyGlyProAsnLeuThrSerSerPheHisValIleGlyGlnIle 290  
DB 1668 GTTAGATCTGTACGTGGGTGTCAGACATCTATCACTTACAGATTGACAAATCACAAG 1727  
QY 291 PheAspGlyValHisPheGlnGlyGlyGlyGlnGlnHisAsnIleGlnThrThrLeu 310  
DB 1728 ATGAAGCTCTTGAATGAAAGATCGACCTTCTCAAAAGATACGACTCTTGAC 1787  
QY 311 IleProAlaGly-----GlyAlaAlaIleThrGlnPheGlyValAspValProGly 327  
DB 1788 GTTCACGTTGGCCAGTGGCTTGGACCATCGTACCGGCAATCAA-----GAACCTAA 1841  
QY 328 AspGlyValLeuValAspHisAlaIlePheArgAlaPheAsnGlyGlyAlaLeuGlyIle 347  
DB 1842 GATTACATCAATGATGTTGATCTCTAGTTTGAAGCGTTATCACAACACCGGACTT 1901  
QY 348 LeuGlyValGlnGly 352  
DB 1902 CTCGCTACGAGGA 1916

RESULT 30  
US-08-746-111-4  
Sequence 4, Application US/08746111  
Patent No. 6066778  
GENERAL INFORMATION:  
APPLICANT: Ginsburg, David  
TITLE OF INVENTION: Compositions And Methods For Screening  
TITLE OF INVENTION: Compounds For Anticoagulant Activity  
NUMBER OF SEQUENCES: 54  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Medien & Carroll, LLP  
STREET: 220 Montgomery Street, Suite 2200  
CITY: San Francisco  
STATE: California  
COUNTRY: United States of America  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/746,111  
FILING DATE: 06-NOV-1996  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Ingolia, Diane E.  
REGISTRATION NUMBER: 40,027  
REFERENCE/DOCKET NUMBER: DM-02536  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 705-8410  
TELEFAX: (415) 397-8338  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 6585 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double

TOPOLOGY: linear  
MOLECULE TYPE: other nucleic acid  
DESCRIPTION: /desc = "DNA"  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 6..6554  
US-08-746-111-4  
Alignment Scores:  
Pred. No.: 0.00471 Length: 6585  
Score: 130.00 Matches: 70  
Percent Similarity: 34.32% Conservative: 46  
Best Local Similarity: 20.71% Mismatches: 120  
Query Match: 4.95% Indels: 102  
Gaps: 16  
US-10-088-045-2 (1-502) x US-08-746-111-4 (1-6585)  
QY 63 AspAlaIleValThrHisAlaProGlnValProProValAspArg----- 78  
DB 1074 GAGGTGATTTGGAAGTATGACCCCGTATGCTCGCAATATGACAAATTTACAGGTCT 1133  
QY 79 -----AspHisProAlaGlyValValValys 87  
DB 1134 CAGCAGCTGGATATTTCTCAACCAAAATTGAAACATTACAGAAAGTTATCTACAGG 1193  
QY 88 MetGlnThrValGlnGlyValMetArgLeuAlaAspGlyValGlnGlnPheThr 107  
DB 1194 CAATATGAAAGAAAGACCTTACCAACGACCTGACAAACCCAGCATCAA----- 1244  
QY 108 PheGlyGlyGlnValProGlyGlnMetIleArgValArgGlnGlyAspThrIleGlnVal 127  
DB 1245 ---CAAGTGGAGTTCTGGCCCTGTTATCAGAGCCAGGTGACAGACACACTCAAGATC 1301  
QY 128 GlnPheSerAsn-----HisProAspSerLysMetProHisAsnValAspPheHis 144  
DB 1302 GTGTCAAAAATATATGCGAGCCGACCTTACAGCATTTACCTCAGCGGTGACCTTCTCT 1361  
QY 145 AlaAlaThrGlyProGlyGlyGlyAlaGlnAlaSerPheThrAlaProGlyHisThrSer 164  
DB 1362 CTTTACGAA-----GATGGAATCAATTCTCTCCACCTCAGGAGCTCACACC-- 1409  
QY 165 ThrPheSerPheLysAlaLeuGlnProGly---LeuTyValTy----- 178  
DB 1410 -----ACGATCAGACAGATTCACACGGGGAACCTTCACTTCAAAATGAGACATTCTA 1463  
QY 179 -----HisCysAlaValAlaPro----- 184  
DB 1464 GAGTTGATGACCACGAAACGATGCCAGTGGCTTACAAAGCCACTACTACAGTGAT 1523  
QY 185 -----ValGlyMetHisIleAlaAsnGlyMetTyGlyLeuIleLeuValGlnProLys 202  
DB 1524 GTGACGTTTCAAGAGGATATGCTCTGGCTGATAGGGGTGCTCTTAATTGTAAGAC 1583  
QY 203 GlnGlyLeuProLys-----ValAspLysGlnTyTyValMetGln 216  
DB 1584 AGTCCCTGACACGAGGGGTGATACAGAGGTGGACAGATCAGACGACGACGCGGTGTT 1643  
QY 217 GlyAspPhe---TyThrLysGlyLysTyGlyGlnGlnGlyLeuGlnProPhe----- 233  
DB 1644 GCTGTGTTTGCAGCAAGAAAGAGCTGTACATTGAGACAAACATCAAAAGTTCTGTGAG 1703  
QY 234 AspMetGlnGlyValAlaIleArgGlnAspAlaGlnGlyValValPheAsnGlySerValGly 253  
DB 1704 AATCTGTGATGAGTGAAGCTGATGATCCCAAGTTT----- 1739  
QY 254 AlaLeuThrGlyGlnGlnAlaLeuGlyValAlaValGlnThrValArgLeuPheVal 273  
DB 1740 -----TACGAATCAAAATATGACGCTATCAACGCTACAGT 1778  
QY 274 GlyAsnGlyGlyProAsnLeuThrSerSerPheHisValIleGlyGlnIlePheAsp--- 292  
DB 1779 -----CCGAGAGCATTTCCACTGTGAGATTCTGTTGATGAC 1817



QY 386 ProAlaProAla-----AsnLeuGlnGlnGlnLeuValAlaGlyLys 399  
Db 2194 ACAGTCATGGCTTACACGGAATAATGATGATGCTTTAGAACTGAAGATGAAGAGATGAT 2253  
QY 400 AlaThrTyrAsp 403  
Db 2254 GCTGACTATGAT 2265

RESULT 32  
US-08-658-340-1  
Sequence 1, Application US/08658340  
Patent No. 5910576  
Patent No. 5910576 5861489  
GENERAL INFORMATION:  
APPLICANT: Bertina, Rogier  
APPLICANT: Reitsma, Pieter  
TITLE OF INVENTION: A method for diagnosing an increased  
TITLE OF INVENTION: risk for thrombosis or a genetic defect causing  
TITLE OF INVENTION: thrombosis and kit for use with the same.  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Akzo No. 5910576 5861489el Patent Department  
STREET: 1300 Piccard Drive, Suite 206  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/658,340  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/454,353  
FILING DATE: 06-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Gornley, Mary E.  
REGISTRATION NUMBER: 34,409  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 948-7400  
TELEFAX: (301) 948-9751  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 6909 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: both  
TOPOLOGY: unknown  
MOLECULE TYPE: cDNA  
US-08-658-340-1

Alignment Scores:  
Pred. No.: 0.00732 Length: 6909  
Score: 128.50 Matches: 85  
Percent Similarity: 33.02% Conservative: 55  
Best Local Similarity: 20.05% Mismatches: 145  
Query Match: 4.89% Indels: 139  
Gaps: 21

US-10-088-045-2 (1-502) x US-08-658-340-1 (1-6909)

QY 63 AspAlaIleValIThrHisAlaProGluValProProProValAspArg----- 78  
Db 1162 GAAGTCATTTGGGACTATGACCTGTATATACGACGATATATGACAAAAATATACAGGTCT 1221  
QY 79 -----AspHisProAlaLysValValValLys 87  
Db 1222 CAGCATTTGATTAATTTCTCAACCAATTGGAAAAACATTATTAAGAAAGTTATGTACACA 1281  
QY 88 Met-----GlnThrValGluLysValMetArgLeuAlaAsp 99

Db 1282 CAGTACGAAGATGATGCTCTTACCAACCAATACAGTGAATCCCATATGAAA---GAAGAT 1338  
QY 100 GlyValGluTyrGlnThrThrPheThrPheGlyGlyGlnValProGlyGlnMetIleArgVal 119  
Db 1339 GGG-----ATTTGGGCTCTTATATCAAGGCC 1365  
QY 120 ArgGluGlyAspThrIleGluValGlnPheSerAsn-----HisProAspSerLys 136  
Db 1366 CAGTCAGACACACACTCAAAATCGCTTCAAAAATATGGCCGCCGCCCTTATACATT 1425  
QY 137 MetProHisAsnValAspPheHisAlaIleThrGlyProGlyGlyGlnValGlnValSer 156  
Db 1426 TACCTCATGAGTGAAGCTTCTCGCTTATGAAGATGAA-----GTCAACTCTTCT 1476  
QY 157 PheThrAlaProGlyIleThrSerThrPheSerPheLysAlaLeuGlnProGly---Leu 175  
Db 1477 TTGACCTCA---GGCAGGAACAACACCATG---ATCAGAGCAGTTCAACCGGGGAACC 1530  
QY 176 TyrValTyr-----HisCys 180  
Db 1531 TATACCTTATAGTGAACATCTTAGAGTTGATGAACCCACGAAAATGATGCCAGTGC 1590  
QY 181 AlaValAlaPro-----ValGlyMetHisIleAlaAsnGlyMetTyr 194  
Db 1591 TTAAACAAGACCACTACTACAGTACAGTGAAGATCATGAGACATCCCTCTGGCTAATTA 1650  
QY 195 GlyLeuIleLeuValGluProLysGlyLeuProLys-----Val 208  
Db 1651 GAGTACTCTTAACTGTAAGACAGATCCCTGGACAGCGAGGAATACAGAGCGACGA 1710  
QY 209 AspLysGluTyrTyrValMetGlnGlyAspPhe---TyrThrLysGlyLysTyrGlyLys 227  
Db 1711 GACATGGAACAGAGCGCTGCTTGTCTGTTGATGAAGAACAAAGCTGTCCTTGAAG 1770  
QY 228 GlnGlyLeuGlnProPhe---AspMetGluLysAlaIleArgGluAspAlaGluTyr 245  
Db 1771 GACAACATCAACAAGTTTGTGAAATCCGATGAGAGTGAAACGATGATGACCCCAAGTTT 1830  
QY 246 ValValPheAsnGlySerValGlyAlaLeuThrGlyLysAlaLeuLysAlaLysVal 265  
Db 1831 -----TATGAATCAAAATCC 1845  
QY 266 GlyLysThrValArgPheValGlyAsnGlyLysProAsnLeuThrSerSerPheHis 285  
Db 1846 ATGAGCACTATCAATGCTATATG-----CCTAGACATTAATCT 1884  
QY 286 ValIleGlyLysIlePheAsp----- 292  
Db 1885 ACTCTTGATTCGCTTGTATGACATGTCACATGACATTCCTGATGTCGGGACCCAG 1944  
QY 293 -----LysValHisPheGlyGlyLys-----GlyGluAsnHisAsn 305  
Db 1945 AATGAATTTTGAACATCACTTCACTGGGACCTGATCATCTATGAAAAAGAGCAGTGAAG 2004  
QY 306 IleGlnThrThrIleLeuLeuProAlaGlyGlyAlaAlaIleThrGluPheLysValAspVal 325  
Db 2005 GACACCTTGAACCTCTTCCCAATGCGTGAAGATCTGTGACG----- 2046  
QY 326 ProGlyAspTyrValLeuValAspHisAlaIlePheArgAlaPheAsnLysGlyAlaLeu 345  
Db 2047 -----GTCACAATGAT-----AATGTTGAAACTTGG 2073  
QY 346 GlyIleLeuLysValGluGlyGluGluAsnHisGluIleTyrSerHisLysGlnThrAsp 365  
Db 2074 ATGTAACTTCCTCAATTCCTAGTGCAGAAACCAAAAGCTGAAATTCAGAGAT 2133  
QY 366 AlaValTyrLeuProGluGlyAlaProGlnAlaIleAspThrGlnGluAlaProLysThr 385  
Db 2134 GTTAAATGTATCCCAATGATGATGAAGACTCATGAGATTTTGAACCTCCAGAAATCT 2193  
QY 386 ProAlaProAla-----AsnLeuGlnGlnGlnLeuValAlaGlyLys 399



```

US-08-454-353A-1
; Sequence 1, Application US/08454353A
; Patent No. 6518016
; GENERAL INFORMATION:
; APPLICANT: Bertina, Rogier
; APPLICANT: Bertina, Pieter
; TITLE OF INVENTION: A method for diagnosing an increased
; TITLE OF INVENTION: risk for thrombosis or a genetic defect causing
; TITLE OF INVENTION: thrombosis and kit for use with the same.
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Akzo No. 6518016el Patent Department
; STREET: 1300 Piccard Drive, Suite 206
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/454,353A
; FILING DATE: 06-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Gormley, Mary E.
; REGISTRATION NUMBER: 34,409
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 948-7400
; TELEFAX: (301) 948-9751
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6909 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; US-08-454-353A-1
Alignment Scores:
Pred. No.: 0.00732 Length: 6909
Score: 128.50 Matches: 85
Percent Similarity: 33.02% Conservative: 55
Best Local Similarity: 20.05% Mismatches: 145
Query Match: 4.89% Indels: 139
DB: Gaps: 21
US-10-088-045-2 (1-502) x US-08-454-353A-1 (1-6909)
QY AspalalalevalThrhIsAlaProgluValProProValaspArg----- 78
::: ||| ::||| |::| |::| |::|
Db 1162 GAAGCATTTGGGACTATGCACCTGTAAATACCAGGAATATGCAAAAAATACAGGCT 1221
79 -----AsphSPoAlaLyValValVallys 87
::: |::| |::| |::| |::|
Db 1222 CAGCATTGGATAATTTCTCAAACCAATTGGAAAACATTATTAAGAAATTATTCACACA 1281
88 Met-----GIuThrValIGluValMetArgLeuLaaSp 99
::: |::| |::| |::| |::|
QY 1282 CAGTACGAAGATGACTCCTTCACCAAACATACAGTGAAATCCCAATATGAAA--GAAAT 1338
100 GIyValIGluTyrgInPheTrpThrPheGIyGIyInValProGIyGImecIlleArGVal 119
||| |::| |::| |::| |::|
Db 1339 GCG-----ATTtGGTCcTATtATTCAGAAC 1365
120 ArgGIuGIyAspThrIleGIuValGlnPheSerAen-----HiSProAspSerLys 136
::: |::| |::| |::| |::|
Db 1366 CAGGTACAGACACACACTCAAAATCGTGTCAAAAATATATGCGCACGCCCGCTATGCAATT 1425
137 MetPrOHIsaenValAspPheHIsAlaIatHrGIyProGIyGIyGIyAlaGIuLaaSer 156
||||| ||| |||

```

Db 1226 TACCCGATGGAGGACCTTCTCGCCCTTATGAAGATGA-----GTCACCTTTCT 1476  
 Qy 157 PheThAlaPrOgLiYhiSthSerThPhSeSerPheYsaLaLeuGlnPrOgLi---Leu 175  
 Db 1477 TTTCACCTCA---GGCAGGAACAACACCATG---ATCAGACAGTTCAACAGGGGAACC 1530  
 Qy 176 TyrValYyr-----HisCys 180  
 Db 1531 TATCTCTTATAGTGGAAACATCTTAGAGTTTGATGAACCCACAGAAAATGATGCCCACTGC 1590  
 Qy 181 AlaValAlaPro-----ValGlyMetHisIleAlaenglyMetTyr 194  
 Db 1591 TTAAACAAGACCAATCTACAGTGAACGTGCATCATCATGAGACATCGCTCGGGCTATATA 1650  
 Qy 195 GlyLeuIleLeuValGluPrOlyGlyGlyLeuProlys-----Val 208  
 Db 1651 GGACTCTCTTAACTCTTAAGTGAAGCAGATCCCTGGACGGGAGATACAGAGGCAGCA 1710  
 Qy 209 AspArgGluYyrTyrValMetGlnGlyAspPhe---TyrThrIleGlyIleGlyGly 227  
 Db 1711 GACATCGAACAACGAGCGCTGTGTGCTGTGTGTTGATGAACAACAACTGTACTTGAG 1770  
 Qy 228 GlnGlyLeuGlnPrOpe-----AspMetGlyIleValIleAspGluAspAlaGluTyr 245  
 Db 1771 GACAACATCAACAAGTTTGTAAGAAATCGATGAGGTGAAGAGTATGACCCCAAGTTT 1830  
 Qy 246 ValValPheasnGlySerValGlyAlaLeuThrGlyIleAsnAlaLeuValAlaVal 265  
 Db 1831 -----TATGAATCAACATC 1845  
 Qy 266 GlyIleYthrValArgIlePheValGlyAsnGlyGlyProAsnLeuThrSerSerPheHis 285  
 Db 1846 ATGAGCATCAATCAATGGCTATAGTG-----CCTGAGACATTAAT 1884  
 Qy 286 ValIleGlyGluIlePheAsp----- 292  
 Db 1885 ACCTTGATGATTCGCTTGTATGACACCTGTCACAGTGCACCTTCTGTAGTGGGAGCCGAG 1944  
 Qy 293 -----LysValHisPheGluGlyGlyLys-----GlyIleAsnHisAsn 305  
 Db 1945 AATGAATTTTGAACATCACTCACTTCAGGGCAGCTATCTATGAGAAAGAGCATGAG 2004  
 Qy 306 IleGlnThrIleLeuIleProAlaGlyGlyAlaIleIleThrGluPheIleValAspVal 325  
 Db 2005 GACACCTTGACCTTCTCCCATGCGTGAGAAATCTGTGACG----- 2046  
 Qy 326 ProGlyAspYrYrValIleuValAspHisAlaIlePheArgAlaPheAsnIleGlyAlaLeu 345  
 Db 2047 -----GTCACAAATGAT-----AATGTTGGAACCTGG 2073  
 Qy 346 GlyIleLeuIleValGluGlyGluIleAsnHisGluIleTyrSerHisIleGlnIleAsp 365  
 Db 2074 ATGTTAACTTCAATGATTTCTAGTCCAAAGACAAAGAAAGCTGAGAGCTGAATTCAGGAT 2133  
 Qy 366 AlaValIleYrLeuPrOgLiGlyAlaProGlnAlaIleAspThrGlnGlyAlaProlysThr 385  
 Db 2134 GTTAAATGATATCCCAATGATGATGAAGACTCATATGAGATTTTGAACCTCCGAATCT 2193  
 Qy 386 ProAlaProAla-----AsnLeuGlnGluIleIleValIleGlyLys 399  
 Db 2194 ACAAGTATGGCTACAGAGAAATGATATCGTTTGAACCTGAAGATGAAGAGATGAT 2253  
 Qy 400 AlaThrTyrAsp 403  
 Db 2254 GCTGACTATGAT 2265  
 RESULT 35  
 US-09-165-019-1  
 ; Sequence 1, Application US/09165019  
 ; Patent No. 6558913  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Bertina, Rogier  
 ; APPLICANT: Reitsma, Pieter





Score:	125.50	Matches:	68
Percent Similarity:	33.98%	Conservative:	37
Best Local Similarity:	22.01%	Mismatches:	89
Query Match:	4.78%	Indels:	115
DB:	4	Gaps:	15

US-10-088-045-2 (1-502) X US-09-252-991A-6517 (1-933)

OY	187	MeChI1s1a1aAaNsG1yMeTyrG1yLys1a1eValG1uProLysG1uLysPro	206
DB	313	ATGAACATGCGCCAGAGAAAGTACGGGGCCATCTTGGC-----	351
OY	207	LysValAspLysG1uTyrTyrValMetG1uLysAspPheTyrThrLysG1uLysTyrG1y	226
DB	352	-----AAGTACGGCCGCATGCCATGAGAAAGTGGCCAAAGACGACGCGC	399
OY	227	GIuG1uG1yLeuG1uInProPheAspMetG1uLysAla1aLeuG1uAspAlaG1uTyrVal	246
DB	400	-----ATGAACATGGCGACGGCGCATGTGGCCACTTGC	435
OY	247	---ValPheAsnG1ySer-----ValG1yAlaLeuThrG1uLysAsn1a1eLysAla	263
DB	436	TGCATGTGCCACGCGTTCCAGCGCCAAAGGGCGCCCTC-----	471
OY	264	LysValG1yG1uThrValArgLeuPheValG1yAsnG1yG1uProAsnLeuThrSerSer	283
DB	472	-----GGCTTCCCAACCTGGCGGCAAC	495
OY	284	PheHisVal11eG1uG1uLeuPheAspLysValHisPheG1uG1yLysG1uLysAsn	305
DB	496	-----GAGTGGCGTGGGGC---GGGACCG	519
OY	304	HisAsn11eG1uThrThrLeuLeuProAlaG1yAla1a1a1eThrG1uPheLysVal	322
DB	520	CAGAGATGAGACCAACCATCTTC-----GGCGGTGC-----	555
OY	324	AspValProG1yAspTyrValLeuValAspHisAla1a1ePheArgAlaPheAsnLysG1y	343
DB	553	-----CAGCGCATCATGGCGCCTGGGC-----	576
OY	344	AlaLeuG1y11eLeuLysValG1uG1yG1uLysAsnHisG1uLeuTyrSerHisLysG1u	365
DB	577	-----GACATCTTGGCGGAGAGGGGTGAAACCTGCCCGCTACGTCCGACCGAA	630
OY	364	ThrAspAlaValTyrLeuProG1uG1yAlaProG1uAla1a1eAspThrG1uAlaPro	383
DB	631	CTCGCGCGGCGTGAAGCTCCCGAAGCAC-----	666
OY	384	LysThrProAlaProAlaAsnLeuG1uG1u11eLysAlaG1yLysAlaThrTyrAsp	405
DB	661	-----AAGGCGGAGCTCGAGGCGCGGCAACAGATCTTCTCG	696
OY	404	SerAsnCyAla1a1aCyHisG1uInProAspG1yLys-----GlyValPro	418
DB	697	GTCACCTGCTGCGCTGCACGCGCCCGAAGGCAAGGATCCCGCCTGTGGGGCGCGG	756
OY	419	AsnAlaPheProProLeuAlaAsnSerAspTyrLeuAsn1a1aAspHisAlaArgAla	438
DB	757	AACCTGACCAACCC-----GGCGCTTCATCTACGGCTCCAGCTACGCCAGCTCAA	811
OY	439	Ser11eVala1a1aAsnG1yLeuSerG1yLys1a1eThrValAsnG1yAsnG1uTyrG1uSer	455
DB	811	CAAAACATCGCGCATGGCGCCGACAGGGCGAG-----	844
OY	459	ValMetProAla1a1eAla-----LeuSerAspG1uG1u11eAlaAsnVal11eThrTyr	476
DB	841	---ATGCGCGGCGCAGAGCCATCTGGGCAAGAAAGGATGACATCTGGCGCGCTAC	897
OY	477	ThrLeuAsnSerPheG1yAsnLysG1y	485
DB	898	ATCTACAACTGTGCGACAAACAGGCG	924

```

US-09-324-867-1
; Sequence 1, Application US/09324867A
; Patent No. 6251632
; GENERAL INFORMATION:
; APPLICANT: Lillcrap, David
; APPLICANT: Cameron, Cherie
; APPLICANT: No. 6251632ley, Colleen
; APPLICANT: Horrocks, L. Suzanne Hoyle
; APPLICANT: Hough, Christine
; TITLE OF INVENTION: Canine Factor VIII Gene, Protein and Methods of Use
; FILE REFERENCE: 1669, 0010002/JAG/BJD
; CURRENT APPLICATION NUMBER: US/09/324,867A
; CURRENT FILING DATE: 1999-06-03
; EARLIER APPLICATION NUMBER: 09/035,141
; EARLIER FILING DATE: 1998-03-059
; EARLIER APPLICATION NUMBER: 60/039,953
; EARLIER FILING DATE: 1997-03-06
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;
; LENGTH: 7032
; TYPE: DNA
; ORGANISM: Canis familiaris
; FEATURES:
; NAME/KEY: CDS
; LOCATION: 1..7029
;
US-09-324-867-1

```

Alignment Scores:	
Pred. No.:	0.0528
Length:	7038
Matches:	97
Conservative:	65
Percent Similarity:	35.14%
Mismatches:	176
Best Local Similarity:	25.10%
Query Match:	4.59%
Indels:	123
Gaps:	25
DB:	3

US-10-088-045-2 (1-502) X US-09-324-867-1 (1-7032

QY	68	h1e1a1aProGluVal1ProProProValAspArgAspHis-----	80
Db	1222	TATGCTCCCTCAGGCCCCCAGCCCCCAATGAATGAAGATGATAAAAATCTGTATTGGAACAAT	1281
QY	81	---ProAlaLysValValValValLysMetGluThrVal-----GluLys	93
Db	1282	GGCTCTCAGCGGAGTGGTGAAGAAAGTACAAAAGAAAGTCCGATTTGGCGCATCAGACAGATGAG	1341
QY	94	ValMetArgLeuAlaSerGluValGluTyrGlnPheThrPheGluLysGluValPro	113
Db	1342	ACATTTAAAGACTCGTGAAAGCATATTCAGATAGAAATCAGAAATCCGAGACCTTACTTAT	1401
QY	114	GlyLysMetL1ealArgValAlaArgGluLysArgThrL1eGluValGlnPheSerAsn-----	131
Db	1402	GGAGAAGATT-----GGAGACACACTCGCTGATATATTTAAAGAAACAACCC	1446
QY	132	---HisProAspSerLysMetProHisAsnValAspPhe-----HisAlaAla	146
Db	1447	AGCGGGCCATATTAACATCTACCTCCATAGGGATCATTAATGTCACTCCTCTGGACACAGGG	1506
QY	147	ThrGlnProGluGlyGlyAlaGluAlaSerPhe-----ThrAlaProGluLysIleThrSer	164
Db	1507	AGATTGCCAAAGAGTGGAAACAATTGAAAGATATGCCAATTCGCGCGGA-----GAG	1560
QY	165	ThrPheSerPheLys-----AlaLeuGlnProGly-----	174
Db	1561	ATATTCAAGTATTAATGGACAGTAGCCGATGAAGATGAGACCAATAATAGATCTCTGG	1620
QY	175	---LeuTyrValLysHisCysAlaValAlaProValGluMetHisIleAlaSerGlyMet	193
Db	1621	TGCGCTGACCCGATTTACTCAAGCTTCATTAATCTGGAGAGAGATCTTAAGCTTCAGACTC	1680
QY	194	TyrGluLeuIleLeuValGlnProLysGluLysLeuProLys-----	207
Db	1681	ATTGAGCCCTCTTTCATCTGCTACAAAGAACTGTATGAATCAAGAGAAACAGATGATG	1740

[illegible]

; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR APPLICATION NUMBER: US 66/094,190  
 ; PRIOR FILING DATE: 1998-07-27  
 ; NUMBER OF SEQ ID NOS: 33142  
 ; SEQ ID NO 8062  
 ; LENGTH: 1944  
 ; TYPE: DNA  
 ; ORGANISM: *Pseudomonas aeruginosa*  
 US-09-252-991A-8062

Alignment Scores:  
 Pred. No.: 0.0102 Length: 1944  
 Score: 119.00 Matches: 107  
 Percent Similarity: 30.83% Conservative: 57  
 Best Local Similarity: 20.11% Mismatches: 175  
 Query Match: 4.53% Indels: 193  
 DB: 4 Gaps: 23

US-10-088-045-2 (1-502) x US-09-252-991A-8062 (1-1944)

QY	27	AlAspLyValAlaGlnProLysSerSerThrValAspAlaAlaAlaLysThrAlaAsn	46
DB	1789	GCCACCCAGGAAGCGCCCTCGCCAGAGATCGATAGCGTTG	1748
QY	47	AlAspAsnAlaAlaSerGlnGlnHisGlnGlyLeuProValIleAspAlaIleVal	66
DB	1747	GGCAGAGGCGACAGACGCCCTGGGTAAAGCCAGCAGCAGCAGCCGCTCTC	1697
QY	67	ThrHisAlaProGluValProProProValAspArgAspHisProAlaLysValVal	86
DB	1696	GTCACGCCCCAGTCGAT	1661
QY	86		86
DB	1660	CTGGCAATAGACATTGTTGTCCTGCTGGGTGCGGCTGAATCTGTGCGCCGCGACAG	1601
QY	87		100
DB	1600	CATCGGGGTGCTCGGACAGACAGCAGAGGTGGAAGATTGGCATCTGGCGACGCG	1541
QY	101	ValGluTrpGlnPheTrpThrPheGlyGlyValn	116
DB	1540	CAGGGCGCGATCTCCGGGTGTGGTGGGGCCCTCGACGCCGTGGTTCCAGGACAGGTT	1481
QY	117	Ile	117
DB	1480	GTGATCGCTGCGCGGGTGTCTGCGCGTGGCTCGTGTGCTGTGATCGTAGGA	1421
QY	118	ArgValArgGluGlyAspThrIle	127
DB	1420	AACCAACGTGCGCAGGTGAACCCGTGTGGGCGGTGACGAAAGTTCAACCGACGGTAGG	1361
QY	128		130
DB	1360	CCGCCGTCACGCTGTGTCGTAGAGATGCGCGAGGCGGTAGGCGTCCGCCAGTTCCGG	1301
QY	131	AsnHis-ProAspSerLysMetProHisAsnValAspPheHisAlaIleThr	147
DB	1300	CAGCATGCGCGTGTGCGCGCGCA	1247
QY	148	GlyProGlyGlyValAlaGluAlaSerPheThrLysProGlyHisThrSe	164
DB	1246	CGTTCCATTCGCGCCACCGGCGGGAAGCGCTTACCTGTGACCGCGGG	1195
QY	164	rThrPheSerPheLysAlaLeuGlnProGlyLeuTrpValIleHisCysAlaValAlaPro	184
DB	1194		1181
QY	184	oValGlyMetHisIleAlaAsnGlyMetTrpGlyLeuIleValAlaGluProGlyGluGly	204
DB	1180	GTTCGGCGATCAGCTTTCCTCTCTACGACACCGGGTCTCGCGCGCAGCGACGAGAAAGC	1121
QY	204	YLeuProLysValAspLysGluTrpValMetGlnGlyAspPheTrpThrLysGlyLys	224

```

Db      1120  CGTGGCCCTCGTCAAA-----ACCGTCCGAT 1094
               |||
Qy      224  sTyrGlyGlnGlnGlnLeuGlnProPheAspMetGluLysAlaIleArgLysAspAlaI 244
               |||
Db      1093  GCGCGCCGAGATGTCGCCG-----GTCGACGCGGAGCCGT 1055
               |||
Qy      244  uTyrValValPheAsnGlySerValGlyAlaLeuThrGlyGluAsnAlaLeuLysAlaI 264
               |||
Db      1054  CCACGCCCATTTCCGTGGCCACGAGGACGCGGACGTCGTCACCATCTGCACACGCGAG 995
               |||
Qy      264  sValGlyGlnThrValArgLeuPheValGlyAsnGlyGlyProAsnLeuThrSerSerPh 284
               |||
Db      994  GGTGGCTGAG---GTCAAGGGGTGTGGCGGTCGCGAATCTGATAGACGCGCGCT 938
               |||
Qy      284  e-----HisValIleGlyGluIlePheAspLysValHisPheGlnGlyGly 300
               |||
Db      937  GGTGGCGCATCAGCGGATGATGACGAGCGGTGTCGATGCGCGCATGCACAGGCT--- 883
               |||
Qy      300  sGlyGluAsnHisAsnIleGlnThrThrLeuIleProAlaGlyGlyAlaAlaIleThrG 320
               |||
Db      882  -----CGACCCCACTCTGTCCTCCGCGGTGCTGTGAGACA 842
               |||
Qy      320  uPheLysValAspValProGlyAspTyrValLeuValAspHisAlaIlePheArgAlaPh 340
               |||
Db      841  CGTCGAGATCAGCTCCAGTCGCGCGTCGAGAGTGGCGGACCATCTCTTGAACTCGT 782
               |||
Qy      340  eAsnLysGlyAlaLeuGlyIleLeuLysValGlnGlyGluAsnHisGlnIleTyrSe 360
               |||
Db      781  TGACCTGGCCGCGCGAGTGGCGGTGCGGGCGGAGAAAGC----- 736
               |||
Qy      360  rHisLysGlnThrAspAlaValTyrLeuProGluGlyAla----- 373
               |||
Db      735  -----GATCTGTGTGAGCCCGACAGTGTCTCATGCCGTTTCCAGCAGT 689
               |||
Qy      374  -----ProGlnAlaIleAspThrGlnGlnAlaProLys---ThrProAlaProAl 389
               |||
Db      688  GCTTGTCTGCACGACGCGGTGATCGGACGACGCTGACGCTGACCCCGAGCGCG 629
               |||
Qy      389  aAsnLeuGlnGlnGlnIleLysAlaGlyLysAlaThrTyrAspSerSerLysAlaIa 409
               |||
Db      628  GGATGTGCGCGCAGCGCTC-----GGCGT 605
               |||
Qy      409  sHisGlnProAspGlyLysGlyValProAsnAlaPheProLeuAlaAsnSerAspTyr 429
               |||
Db      604  TCATCAGACC---GCCAAGTCCCGGACCGCTCCGGACCTGTGATGGC----- 553
               |||
Qy      429  rLeuAsnAlaAspHisAlaArgAlaAlaSerIleValAlaAsnGlyLeuSerGlyLys 449
               |||
Db      552  -----CATGCTGAGCGCGCGCAGGTG-----GGCTTCGTAGATCAGCG 515
               |||
Qy      449  eThrValAsnGlyAsnGlnTyrGlnSerValMet 460
               |||
Db      514  TCGCGTCCAGGAGCGGTACCGCGCGCGCTC 481
               |||

```

RESULT 39  
US-09-252-991A-6506  
Sequence 6506, Application US/09252991A

Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 6506  
LENGTH: 477  
TYPE: DNA

TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-6506

Alignment Scores:  
Pred. No.: 0.00128 Length: 477  
Score: 118.50 Matches: 44  
Percent Similarity: 37.23% Conservative: 26  
Best Local Similarity: 23.40% Mismatches: 57  
Query Match: 4.51% Indels: 61  
Gaps: 8

US-10-088-045-2 (1-502) x US-09-252-991A-6506 (1-477)

```

Qy      305  AsnIleGlnThrThrLeuIleProAlaGlyGlyAlaAlaIleThrGlnPheLysValAsp 324
               |||
Db      8  AGCATCGACACCACTCTC-----GCCGTGC----- 37
               |||
Qy      325  ValProGlyAspTyrValLeuValAspHisAlaIlePheArgAlaPheAsnLysGlyAla 344
               |||
Db      38  -----CACGCGATCATGCGCGCTGCGGCGC----- 61
               |||
Qy      345  LeuGlyIleLeuLysValGlnGlyGluAsnHisGlnIleTyrSerHisLysGlnThr 364
               |||
Db      62  ---GACATCTCTGCGGACGAGAGCGGTGAAGAACTGCGCGCTGACGTCGACCGAACTC 118
               |||
Qy      365  AspAlaValTyrLeuProGlnGlyAlaProGlnAlaIleAspThrGlnGlnAlaProLys 384
               |||
Db      119  GCCGCGCTGAGCTGCGGAGGACCC----- 145
               |||
Qy      385  ThrProAlaProAlaAsnLeuGlnGlnIleLysAlaGlyLysAlaThrTyrAspSer 404
               |||
Db      146  -----AAGCGCAGCTGACGAGCGCGGACGACGATCTCTCGCTC 184
               |||
Qy      405  AsnGlyAlaAlaCysHisGlnProAspGlyLys-----GlyValProAsn 419
               |||
Db      185  AACCTGCTGCGCTGCGGACCGCGGAGGACGAGTACCGCTGTCGCGCGCGGAC 244
               |||
Qy      420  AlaPheProProLeuAlaAsnSerAspTyrLeuAsnAlaAspHisAlaArgAlaIaSer 439
               |||
Db      245  CTGACCAACCC-----GGCGGTTCATCTACGCTCCACGTACGCCCACTCCAAACA 298
               |||
Qy      440  IleValAlaAsnGlyLeuSerGlyLysIleThrValAsnGlyAsnGlnTyrGlnSerVal 459
               |||
Db      299  ACCATCGGCATGCGCGGACGCGGCGCAG----- 325
               |||
Qy      460  MetProAlaIleAla-----LeuSerAspGlnGlnIleAlaAsnValIleThrTyrThr 477
               |||
Db      326  ATGCCGCGCCAGAGGACCATCTGCGGCAAGAAAGTGCACATCTGCGCGCTACATC 385
               |||
Qy      478  LeuAsnSerPheGlyAsnLysGly 485
               |||
Db      386  TACACCTGTGCGCAACCAAGGC 409
               |||

```

RESULT 40

US-09-252-991A-1456/c  
Sequence 1456, Application US/09252991A

Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 1456  
LENGTH: 1494  
TYPE: DNA

ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-1456

## Alignment Scores:

Pred. No.:	0.00762	Length:	1494
Score:	118.50	Matches:	77
Percent Similarity:	35.78%	Conservative:	45
Best Local Similarity:	22.58%	Mismatches:	138
Query Match:	4.51%	Indels:	83
DB:	4	Gaps:	16

US-10-088-045-2 (1-502) x US-09-252-991A-1456 (1-1494)

```

QY 18 LeuMetLeuSerGlyCysSerAsnGlnAlaAspLysAlaAlaGlnProLys----- 34
DB 1385 TTACCCGTAGACAAAGTGCTCG-----GCGGCTCGCCGGCGCTGGCGGTGG 1341
QY 35 SerSerThrValAspAlaAlaAlaLysThrAlaAsnAlaAspAsnAlaAlaSerGlnGlu 54
DB 1340 TCGGCTCGCGCGCGCGCGCGCGCTGCGCTG- GCGGCTCGCGAGGTGGCGCAGAG 1282
QY 55 HisGlnGlyGluLeuProValIleAspAlaIleValThrHisAlaProGluValProPro 74
DB 1281 TACGACTACGAAGTCTG-----ATCGCCGCG 1258
QY 75 ProValAspArgAspHisProAlaLysValValLysMetGluThrValGluLysVal 94
DB 1257 CCGCTGAC----- 1249
QY 95 MetArgLeuAlaAspGlyValGluThrGlnPheThrPheGlyGlyGlnValProGly 114
DB 1248 CTGGAGATGCTCGCGGCTTTCAGACATCGCGCGCTGCGCTATGCGCGCAAGTCCCGCGC 1189
QY 115 GlnMetIleArgValArgGluGlyAspThrIleGluValGlnPheSerAsnHisProAsp 134
DB 1188 GTGAGCTTGGCGCGCAAGCGAGGCTGGGTGGCTGCGGTGGCTTACCAACAGCGTGGAC 1129
QY 135 SerLysMetProHisAsnValAspPheHisAlaAlaThrGlyProGlyGlyAlaGlu 154
DB 1128 GAA-----CCGACCACCATCCACTGACACCGCGCTGCGCATCGAGATGACGCGC 1075
QY 155 AlaSerPheThrAla-----ProGlyHisThrSerThrPheSerPheLys 169
DB 1074 GTGCGGTACATTTCGCGCGCGCGCGCGCGCGCGCGAGGCTTCATCACTACAGTTCAAG 1015
QY 170 AlaLeuGlnProGlyLeuThrValTyrHisCysAlaValAlaProValGlyMetHisIle 189
DB 1014 ACCCAGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 958
QY 190 AlaAsnGlyMetTyrGlyLeuIleLeuValGluProLysGluGlyLeu----- 205
DB 957 GCGCGCGGACTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 898
QY 206 ProLysVal-----AspLysGluTyrTyrValMetGlnGlyAspPheTyrThrLysGly 223
DB 897 GAGAAAGTTCTCTGCTGAAGACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 865
QY 224 LysTyrGlyGluGlnGly---LeuGlnProPheAspMet---GluLysAlaIleArgGlu 241
DB 864 -----GACGAGAGGGCGCGCTTACCCCTTCAGGTGCGCGCGCGCGCGCGCGTGA 811
QY 242 Asp-----AlaGluTyrValValPheAsnGlySer----- 251
DB 810 GGCACCGCGCGCGCGCTATTCGACGATCAACGCAACGATGCGCGCGCATCGACCTGCGCG 751
QY 252 ValGlyAlaLeuThrGlyGluAsnAlaLeuLysAlaLysValGlyGluThrValArgLeu 271
DB 750 GCGCGCGAGATCGTCCGGGTGCGCTGGAAGTGAACGACACACGCTACCTACGCTGC 691
QY 272 PheValGlyAsnGlyGlyProAsnLeuThrSerSerPheHisValIleGlyGluIlePhe 291
DB 690 AACCTGCCCAAGCGCGAGGCGAGATC-----TATGCCATCGATGCGCATCGCGGTG 640

```

```

QY 292 AspLysValHisPheGluGly-----GlyLysGlyGluAsnHisAsnIleGln 307
DB 639 GAACACCGCGCGCTTCGAGGCGCAGTACGATCGGTCCCGGATGCGCTGGAGCTGGCG 580
QY 308 ThrThrLeuIleProAlaGlyValAlaAlaIleThrGluPheLysValAspValProGly 327
DB 579 CTGAAGGTGCGGAAGCGGCGAC-CGAGCTGTCTCGCGCGAGCGCGCGGTACGCTGGC 521
QY 328 Asp 328
DB 520 GAC 518

```

Search completed: September 7, 2003, 08:08:11  
Job time: 27346 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: September 7, 2003, 00:25:01 / Search time 375 Seconds  
(without alignments)  
3076.903 Million cell updates/sec

Title: US-10-088-045-2  
Perfect score: 2626  
Sequence: 1 MSKPTLIKTTILICALSALML.....NKGQUSADVAKAKTKKEN 502

Scoring table: BLOSUM62  
Xgapop 10.0, Ygapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Delop 6.0, Delext 7.0

Searched: 1537136 segs, 114998732 residues  
Total number of hits satisfying chosen parameters: 3074272

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODE=frame+ p2n.model -DEV=xlh  
-O=/cg2\_1/USFTO.spool/US10088045/rnat\_04092003\_083147\_4078/app\_query.fasta\_1.647  
-DB=Published.Applications\_NA -QFMT=fastcap -SUFFIX=rmpb -MINMATCH=0.1  
-LOOPEXT=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62  
-TRANS=human40.cci -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=40 -MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=200000000 -USER=US10088045 @CGN 1 1 244 @rnat\_04092003\_083147\_4078  
-NCPU=6 -ICPU=3 -NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSRBLCK=100  
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FAPOP=6 -FAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:  
1: /cg2\_6/prodata/1/pubpna/US07\_PUBCOMB.seq:\*  
2: /cg2\_6/prodata/1/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cg2\_6/prodata/1/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cg2\_6/prodata/1/pubpna/US06\_PUBCOMB.seq:\*  
5: /cg2\_6/prodata/1/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cg2\_6/prodata/1/pubpna/PCTUS\_PUBCOMB.seq:\*  
7: /cg2\_6/prodata/1/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cg2\_6/prodata/1/pubpna/US08\_PUBCOMB.seq:\*  
9: /cg2\_6/prodata/1/pubpna/US09\_PUBCOMB.seq:\*  
10: /cg2\_6/prodata/1/pubpna/US09C\_PUBCOMB.seq:\*  
11: /cg2\_6/prodata/1/pubpna/US09C\_PUBCOMB.seq:\*  
12: /cg2\_6/prodata/1/pubpna/US10\_NEW\_PUB.seq:\*  
13: /cg2\_6/prodata/1/pubpna/US10\_PUBCOMB.seq:\*  
14: /cg2\_6/prodata/1/pubpna/US10\_PUBCOMB.seq:\*  
15: /cg2\_6/prodata/1/pubpna/US10\_NEW\_PUB.seq:\*  
16: /cg2\_6/prodata/1/pubpna/US60\_NEW\_PUB.seq:\*  
17: /cg2\_6/prodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	152.5	5.8	1125	14	US-10-127-032-31 Sequence 31, Appl

2	142	5.4	1533	10	US-09-738-626-1053	Sequence 1053, Ap
3	142	5.4	3309400	10	US-09-738-626-1	Sequence 1, Appl
4	140.5	5.4	1479	10	US-09-738-626-3280	Sequence 3280, Ap
5	139	5.3	1722	10	US-09-938-842A-1675	Sequence 1675, Ap
6	136.5	5.2	980	15	US-10-174-693-208	Sequence 208, Ap
7	132	5.0	1713	10	US-09-938-842A-1910	Sequence 1910, Ap
8	129	4.9	2230	12	US-10-361-460-78	Sequence 78, Appl
9	128.5	4.9	2297	14	US-10-115-563-17	Sequence 17, Appl
10	128.5	4.9	2297	14	US-10-115-563-27	Sequence 27, Appl
11	128.5	4.9	2297	14	US-10-115-563-28	Sequence 28, Appl
12	128.5	4.9	6909	10	US-09-880-107-2275	Sequence 2275, Ap
13	128.5	4.9	6925	14	US-10-115-563-13	Sequence 13, Appl
14	128.5	4.9	6925	14	US-10-115-563-26	Sequence 26, Appl
15	125	4.8	9025608	14	US-10-156-761-1	Sequence 1, Appl
16	119	4.5	1626	14	US-09-938-842A-1754	Sequence 1754, Ap
17	118.5	4.5	9025608	14	US-10-156-761-1	Sequence 1, Appl
18	117	4.5	2151	14	US-10-127-032-72	Sequence 72, Appl
19	117	4.5	8911	13	US-10-044-090-17	Sequence 17, Appl
20	114.5	4.4	1791	9	US-09-338-723A-5	Sequence 5, Appl
21	114.5	4.4	1791	13	US-10-080-233-1	Sequence 1, Appl
22	114.5	4.4	1791	13	US-10-080-233-1	Sequence 1, Appl
23	114.5	4.4	13842	10	US-09-861-289-30	Sequence 30, Appl
24	114.5	4.4	13842	10	US-09-860-846-30	Sequence 30, Appl
25	114.5	4.4	13842	11	US-09-988-384B-30	Sequence 30, Appl
26	114.5	4.4	13842	11	US-09-836-821-30	Sequence 30, Appl
27	114.5	4.4	36778	10	US-09-861-289-5	Sequence 5, Appl
28	114.5	4.4	36778	10	US-09-860-846-5	Sequence 5, Appl
29	114.5	4.4	36778	11	US-09-836-821-5	Sequence 5, Appl
30	114.5	4.4	37948	11	US-09-988-384B-5	Sequence 5, Appl
31	114.5	4.4	38506	11	US-09-793-708-19	Sequence 19, Appl
32	114.5	4.4	38506	12	US-10-201-365-1	Sequence 1, Appl
33	114.5	4.4	38506	12	US-10-160-539-19	Sequence 19, Appl
34	114.5	4.3	1938	14	US-10-156-761-4392	Sequence 4392, Ap
35	114	4.3	2670	13	US-10-002-600-44	Sequence 44, Appl
36	113.5	4.3	6026	10	US-09-995-749A-1	Sequence 1, Appl
37	113	4.3	7914	13	US-10-095-718-3	Sequence 3, Appl
38	113	4.3	27360	10	US-09-070-927A-164	Sequence 164, App
39	112.5	4.3	2464	12	US-09-960-706-639	Sequence 639, App
40	112.5	4.3	2464	14	US-10-205-823-290	Sequence 290, App
41	112.5	4.3	2481	9	US-09-853-386-89	Sequence 89, Appl
42	112.5	4.3	2481	12	US-10-101-510-139	Sequence 139, App
43	112.5	4.3	2481	14	US-10-211-239-8	Sequence 8, Appl
44	112.5	4.3	2481	14	US-10-288-222A-5	Sequence 5, Appl
45	112	4.3	1608	14	US-10-156-761-555	Sequence 555, App

ALIGNMENTS

RESULT 1  
US-10-127-032-31  
Sequence 31, Application US/10127032  
Publication No. US20030113742A1  
GENERAL INFORMATION:  
APPLICANT: Whiteley, Marvin  
APPLICANT: Banger, M. Gita  
APPLICANT: Lory, Stephen  
TITLE OF INVENTION: Greenberg Everett Peter  
TITLE OF INVENTION: BIOFILM FORMATION  
FILE REFERENCE: UZ-070CP  
CURRENT APPLICATION NUMBER: US/10/127,032  
CURRENT FILING DATE: 2002-04-19  
PRIOR APPLICATION NUMBER: US 60/285,190  
PRIOR FILING DATE: 2001-04-20  
PRIOR APPLICATION NUMBER: US 60/344,142  
PRIOR FILING DATE: 2001-10-24  
NUMBER OF SEQ ID NOS: 170  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 31  
LENGTH: 1125  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-10-127-032-31





```

QY      84 ValValValValMetGluThrValGluValValMetArgLeuAlaSerGlyValGluTyr 103
      |||
      |||
      |||
Db      175 GTGCACCTTTCCTGGAGCTCAGACTGGGAGAGTCAAGTTTGGCCGATGTCACACG 234
QY      104 GlnPheThrPheGluGluGluValProGlyGlnMetLeuArgValArgGluGlyAsp 123
      |||
      |||
      |||
Db      235 AAGACGTGGGGTTTCAATGACACTATTGGGCGCAGCTTGTTGGTGAAGAAAGGTAT 294
QY      124 ThrLeuValGlnPheSerAsnHisProAspSer----- 135
      |||
      |||
      |||
Db      295 GAGCTCCACCTTATGATGATAAACAAATTGGATGAATGACCACTGGCATGGCATGGC 354
QY      136 ---LysMetProHisAsnValAspPhe-----HisAlaAlaThrGlyProGlyGly 152
      |||
      |||
      |||
Db      355 ATGAAGTTCGGCGGATTCGTGATGGTGTCCCGCACTACCCGATCGGGCTGGGAGACG 414
QY      153 AlaGluAlaSerPheThrAlaProGlyHisThrSerThrPheSerPheLysAlaLeuGln 172
      |||
      |||
      |||
Db      415 TGGTCAACCAAGCTGCTGGCCAAATGATGACGCCCTTGTGTATCCACCGGCACACT 474
QY      173 ProGlyLeuTyrValTyrHisCysAlaValAlaProValGlyMetHisLeuAlaSerGly 192
      |||
      |||
      |||
Db      475 CATTGGCTG-----ACAGTTTGCATGGCTACCGGTGT 507
QY      193 MetTyrGlyLeuValLeuValGluProLysGluGlyLeuProLysValAsp-----Lys 210
      |||
      |||
      |||
Db      508 TTGGCGGGGATGATCATTTGTGAA---GATGAAGCAACAGACAAGCTGATTCGCCACGC 564
QY      211 GluTyrTyrVal-----MetGlnLysAspPheTyrThrLysGly 223
      |||
      |||
      |||
Db      565 GAGTACGGGTGGACGATATTCGCTGTTTAAATGATGATCAGCGCTTCTTGAAGAAGCGT 624
QY      224 LysTyrGlyLysGlnGlyLeuGlnProPheAspMetGluLysAlaLeuArgGluAspAla 243
      |||
      |||
      |||
Db      625 TCCCTTGATGAGGAACACTCCCC---GATCTTGGGCTGTGGGCGCATACCCCACT 678
QY      244 GluTyrValValPheAsnLysSerValGlyAlaLeuThrGlyLysAlaLeuLysAla 263
      |||
      |||
      |||
Db      679 GCCAATGGCATTTACCAATGGCAGCTTGTATGCCACGCGGGGTTCGGTCCGGGTG 728
QY      264 LysValGlyLysThrValArgLeuPheValGlyAsnGlyLysProAsnLeu----- 280
      |||
      |||
      |||
Db      739 CTCACGCGCTCCAAATGCGGTCTAT-----AACTTGGCGTTTTC 780
QY      281 ---ThrSerSerPheHisValIleGly-----GluIlePheAsp 292
      |||
      |||
      |||
Db      781 GACACCGGACCTTCCAACTCATTTGCCAGCATTCGCGGATTCGCTGGATGACCTCAAGAC 840
QY      293 LysValHisPheGluGluGlyLysGlyLysAsnHisAsnIleGlnThrThrLeuIlePro 312
      |||
      |||
      |||
Db      841 CGCACACACTTGGCTATGGCCCGAGGAGCGGTGGGAATCGTCGTGAGCTAGAGCCC 900
QY      313 AlaGlyGlyAlaAlaIleThr-----GluPheLysValAspValPro 326
      |||
      |||
      |||
Db      901 ---GGCAGAGACGTCACCTTGGATCTGTAGGTTTTCAGGACAACCTACGCGCTCCCT 954
QY      327 GlyAspTyrValLeuValAspHisAlaIlePheArgAlaPheAsnLysGlyAlaLeuGly 346
      |||
      |||
      |||
Db      955 GATGATGAGTTCGTGGCCGATTTCCGCGATCTCAAGTCTCTTCAG----- 999
QY      347 IleLeuLysValGluGlyLysGluAsnHisGluIleTyrSerHisLysGlnThrAspAla 366
      |||
      |||
      |||
Db      1000 CTGCTCACCATCAACGCGCTTCGAT-----GATGCT 1032
QY      367 ValTyrLeuPro-----GluGlyAlaProGlnAlaIleAsp 378
      |||
      |||
      |||
Db      1033 GCGCAAGCACTCTTGGCGGCGCTGCTGTGAATTCACCGAACCTGACGTCACTGAT 1092
QY      379 ThrGlnGlu-AlaProLys-----ThrProAlaProAlaAsnLeuGlnGluGlnIleLys 396
      |||
      |||
      |||
Db      1093 GCCACTGAAAGCACTTCATCATGAACACTTCTCCATCAACGATC----- 1138
QY      396 salaglyLysAlaThrTyrAspSerAsnCysAlaAla 408

```

```

Db      1139 -----TACAGATGACATGACGCGCG 1159
RESULT 3
US-09-738-626-1
; Sequence 1, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OKAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; PRIOR FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 1
; LENGTH: 3309400
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-1
Alignment Scores:
Pred. No.: 0.73 Length: 3309400
Score: 142.00 Matches: 77
Percent Similarity: 34.53% Conservative: 58
Best Local Similarity: 19.69% Mismatches: 180
Query Match: 5.41% Indels: 78
DB: Gaps: 13
US-10-088-045-2 (1-502) x US-09-738-626-1 (1-3309400)
QY      44 ThrAlaAsnAlaAspAsnAlaAlaSerGlnGluHisGlnGlyLeuProValIleAsp 63
      |||
      |||
      |||
Db      3169437 ACAGTTTCTGCTCGCGGCGCTGCTACTCGCGCACCGGGCGCTGG-CCGCTGACCA 3169495
QY      64 AlaIleValThrHisAlaProGluValProProProValAspArgAspHisProAlaLys 83
      |||
      |||
      |||
Db      3169496 GGCACCTGACCCCGCTGCTCG-GCACGAGTCTCTCTTGGCCCCCACTCCACCCCC 3169554
QY      84 ValValValValMetGluThrValGluValValMetArgLeu----- 97
      |||
      |||
      |||
Db      3169555 ACTGGCCTCGGTAGCGCCAGCGGTGGCGCGGACACTACCGCGCGGCTCTCCCTGGAT 3169614
QY      98 AlaAspGlyValGluTyrGlnPheThrPheGlyGlyGlnValProGlyGlnMetIle 117
      |||
      |||
      |||
Db      3169615 ATCGCGGCGATGACCAACCAAGCTGGGGATGCTGTGACACCGGGGATGGCGGCATT 3169674
QY      118 ArgValArgGluGlyAspThrIleGluValGlnPheSerAsnHis---ProAspSerLys 136
      |||
      |||
      |||
Db      3169675 GAGGCCACCGCGCGACGCTCTCCAGCTGATATGACCAATGAATGCTGTAAGAGACC 3169734
QY      137 MetPro-----HisAsnValAspPheHisAlaAlaThrGlyProGlyGlyAlaGlu 154
      |||
      |||
      |||
Db      3169735 TCCATGACATGGATGCGATGCGACATCCACAGCAGCGGAGGTGCGCGGCATGACC 3169794
QY      155 AlaSerPheThrAlaProGlyHisThrSerThrPheSerPheLysAlaLeuGlnProGly 174
      |||
      |||
      |||
Db      3169795 CAGAACCCATTGAACCTGGCGAGTCTTCTCTATGTTTGAAGTCCCCCAAGGTGGC 3169854

```



Db 721 ATGATCCGACATCTCATGACGAGATATCCCGCTGCTACCGGACTTGAGGCT 780  
 Qy 264 LysValAlaGluThrValArgLeu---PheValGlyAsnGlyProAsnLeuThrSer 282  
 Db 781 CGCCGGGCGCAAGAGCGCCGCTGGTTTATCAATCCGGCGTGAACCATC----- 834  
 Qy 283 SerPheHisValIle-----GlyGluIlePheAsnLysValHisPheGluGlyLys 300  
 Db 835 ---TTCAAGGTGGCGCTGGTGTACCGCATGATGTCACCCACAGGAGGCTTCCCT 891  
 Qy 301 GlyGluAsnHisAsnIleGlnThrLeuIleProAlaGlyAlaIleThrGlu 320  
 Db 892 GTCCAGCCCTGGGAGACCGCATGATCTACTGTGATGGCGAG----- 936  
 Qy 321 PheLysValAspValPro-----GlyAspTyrValLeuValAsnHisAlaIlePhe 337  
 Db 937 ---CGTGTGACGTCGAGGTATCTCTGGGAGCGGACATCTCCGCGTCACGGCTTGGCG 993  
 Qy 338 ArgAlaPheAsnLysGlyAlaLeuGlyIleLeuLysValGluGlyGluAsnHisGlu 357  
 Db 994 GTGGGTAAAGACGACCGCGCTTCCGCTGATCCGACCGCGCGGCGGCGCCCGC 1053  
 Qy 358 -----IleTyrSerHisLysGln 363  
 Db 1054 CCGGATTCGACTTCCCGAGTTGTCTCCACCGGACTGCTTGTCTCTCGTAAGCCA 1113  
 Qy 364 ThrAspAlaValTyrLeuProGluGlyAlaPro 374  
 Db 1114 GCAGACCGTGCATCTCTGCGGAGGCAACCA 1146

## RESULT 5

US-09-938-842A-1675  
 ; Sequence 1675, Application US/09938842A  
 ; Patent No. US20020160378A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Harper, Jeff  
 ; APPLICANT: Krepes, Joel  
 ; APPLICANT: Wang, Xun  
 ; APPLICANT: Zhu, Tong  
 ; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING  
 ; FILE REFERENCE: SCRIPI300-3  
 ; CURRENT APPLICATION NUMBER: US/09/938, 842A  
 ; PRIOR FILING DATE: 2001-08-24  
 ; PRIOR APPLICATION NUMBER: US 60/227, 866  
 ; PRIOR FILING DATE: 2000-08-24  
 ; PRIOR APPLICATION NUMBER: US 60/264, 647  
 ; PRIOR FILING DATE: 2001-01-16  
 ; PRIOR APPLICATION NUMBER: US 60/300, 111  
 ; NUMBER OF SEQ ID NOS: 5379  
 ; SEQ ID NO 1675  
 ; LENGTH: 1722  
 ; TYPE: DNA  
 ; ORGANISM: Arabidopsis thaliana  
 ; US-09-938-842A-1675

## Alignment Scores:

Pred. No.: 3e-05 Length: 1722  
 Score: 139.00 Matches: 101  
 Percent Similarity: 33.48% Conservative: 54  
 Best Local Similarity: 21.81% Mismatches: 137  
 Query Match: 5.29% Indels: 171  
 DB: 10 Gaps: 26

US-10-088-045-2 (1-502) x US-09-938-842A-1675 (1-1722)

Qy 80 HisProAlaLysValValIleValMetGluThrValGluLysValMetArgLeuAlaAsp 99  
 Db 49 CACTCGCGCTCGGCGGAGTAGTGAATCAACGTGGAG----- 87  
 Qy 100 GlyValGluTyrGlnPheTrp-----ThrPheGly 109

Db 88 ---GTTGAGTAAAGTACTGCTGCGCGGATTTGAAGAAAGAAATGTTATGCCATCAAC 144  
 Qy 110 GlyGlnValProGluGlnMetIleArgValArgGluGlyAspThrIleGluValGlnPhe 129  
 Db 145 GGCAGATTTCAGAGGCCAAGATGACAGCATGCGCGGAGACAGCATATTATC----- 198  
 Qy 130 SerAsnHisProAsnSerLysMetProHisAsn-----ValAspPheHisAla--- 145  
 Db 199 ---CACGTCTCAACAAATCTCCACGAAAGGTGTGTATTCATTGGACAGGCATATA 252  
 Qy 146 -----AlaThrGlyProGluGlyValAlaGluAlaSerPheThr 158  
 Db 253 CGTCAAGAAAGACTCCATGAGCTGATGAGCAGCAGCATGTGACCCAGTCTTATTAAT 312  
 Qy 159 AlaProGluHisThrSerThrPheSerPheLysAlaLeuGlnProGluLeuTyrValTyr 178  
 Db 313 ---CTGGCAGACTTTCATTACATTAAGTTATTTGATAGGCGGGAACAATTTTAC 369  
 Qy 179 HisCysAlaValAlaProValGlyMetHisIleAlaAsnGlyMetTyrGlyLeuIleLeu 198  
 Db 370 GAT-----GACACATACGGAATGCAAAAGATCATCGGAGCATATACGGAATGTATA 420  
 Qy 199 Val---GluProLysGluGlyLeuProLysValAspLysGluTyrTyrValMetGlnGly 217  
 Db 421 GTAAGATCAACAAAGAGAGATTAA---ATTAGATGAGAAAGTTCAATCTTGCTCAGT 477  
 Qy 218 AspPheTyrThrLysGlyLysTyrGlyGluGlnGlyLeuGlnProPheAspMetGluLys 237  
 Db 478 GACTGTGGGACCAAAAGCATTCACGGCCAA----- 507  
 Qy 238 AlaIleArgLysAspAlaGluTyrValAlaPheAsnGlySerValGlyAlaLeuThrGly 257  
 Db 507 ----- 507  
 Qy 258 GluAsnAlaLeuLysAlaLys-----ValGlyGluThrValArgLeuPheVal 273  
 Db 508 GAATCGCTCTTCT 567  
 Qy 274 ---GlyAsnGlyGlyProAsnLeuThrSerSerPheHisValIleGlyGluIlePheAsp 292  
 Db 568 AATGGAAGAGACAGTTCATTTGTTTCAAGACGCTAT-----TTTAAC 612  
 Qy 293 LysValHisPheGluGlyGlyLys-----GlyGluAsnHisAsnIle 306  
 Db 613 AAA-----GGAGGAGAGAAAGATGTATGCACGTTTAAAGAAATGATCACTGT 660  
 Qy 307 ---GlnThrThrLeuIlePro-----AlaGlyGly 315  
 Db 661 GCACCTCAAACTCTGAGTCGAACCCAAATGAGTGAACCGTCTTCGAAATCGGTAGCACA 720  
 Qy 316 AlaAlaIleThrGluPheLysValAspValProGlyAspTyrValLeuValAsnHisAla 335  
 Db 721 ACTGCTCTGCT 771  
 Qy 336 IlePheArgAlaPheAsnLysGlyAlaLeuGlyIleLeuLysValGluGlyGluAsn 355  
 Db 772 -----GTTGAGCTGACGCGAAC 789  
 Qy 356 His-----GluIleTyrSerHisLysGlnThrAspAla 366  
 Db 790 TACGTGCAACGTTTCAACCGTCACAGCATTTAGAGCTTATTC---GGCGAACTTATTC 846  
 Qy 367 ValTyrLeuProGluGlyAla----- 373  
 Db 847 GTTCTCTTAAACCAACGCACTTCATCAAGAAAGTACTGTCTCGTGGCGTTTCT 906  
 Qy 374 -----ProGlnAlaIleAsp-----ThrGln 380  
 Db 907 GCGCGAAGAACCAAAATCTCTCAAGCACTCAACGCTGATAATTACGTTGATGCCACGTAG 966  
 Qy 381 GluAlaProLysThrProAlaProAlaAsn---LeuGlnGluGlnIleLysAlaGlyLys 399  
 Db 967 TCAGGCCCATCTCATCACCACCGGTGACTCCAACTGGAACGACACAGATCGAGCAAA 1026

Qy	400	AlaThrTyrArgSerAsnGlyAlaIalaCysHisGlnProAspGlyValGlyValProAsn	419
Db	1027	AGCTTCTCGAAGAAAGATCTTCCCGCT	1066
Qy	420	AlaPheProIleValAlaAsnSerAspTyrIleu-AsnAlaAspHisAlaIargAlaIaIase	439
Db	1069	-----CCGCCGAGAAATCATGATGACCGCATTAATCTCTCTCCATCAACACAGA-----	1111
Qy	439	rIleValAlaAsnGlyIleuSerGlyValIleIleThrValAsnGlyAsnGlnTyrGluSerVa	459
Db	1115	-----ATCTTACAGAAAGATTACAGAAATGGTCAATCAACAACGTTCTCA	1155
Qy	459	IleuPhePro 461	
Db	1159	TTATCCG 1165	
RESULT 6			
US-10-174-693-208			
; Sequence 208, Application US/10174693			
; Publication No. US20030131373A1			
GENERAL INFORMATION:			
; APPLICANT: Blockberg, Leonard N.			
; APPLICANT: Havukkala, Ilkka			
; TITLE OF INVENTION: Materials and Methods for the			
; TITLE OF INVENTION: Modification of Plant Lignin Content			
; FILE REFERENCE: 11000.1003c5			
; CURRENT APPLICATION NUMBER: US/10/1174, 693			
; CURRENT FILING DATE: 2002-06-18			
; PRIOR APPLICATION NUMBER: US 08/975, 316			
; PRIOR FILING DATE: 1997-11-21			
; PRIOR APPLICATION NUMBER: US 08/713, 000			
; PRIOR FILING DATE: 1996-09-11			
; PRIOR APPLICATION NUMBER: US 09/169, 789			
; PRIOR FILING DATE: 1998-10-09			
; PRIOR APPLICATION NUMBER: US 09/615,192			
; PRIOR FILING DATE: 2000-07-12			
; NUMBER OF SEQ ID NOS: 407			
; SOFTWARE: FastSeq for Windows Version 3.0			
; SEQ ID NO: 980			
; LENGTH: 980			
; TYPE: DNA			
; ORGANISM: EucaIyptus grandis			
US-10-174-693-208			
Alignment Scores:			
	Pred. No.:	2,52e-05	length: 980
	Score:	136.50	Matches: 71
	Percent Similarity:	34.51%	Conservative: 46
	Best Local Similarity:	20.94%	Mismatches: 129
	Query Match:	5.20%	Indels: 93
	DB:	15	Gaps: 14
US-10-088-045-2 (1-502) x US-10-174-693-208 (1-980)			
Qy	78	ArgAspHisIleProAlaIalaValAlaValIuYmeGlnThrValGluIuYValIleArgIleu	97
Db	27	CGTAACACTACACATTCATGTTGGTAAATGAGAATATCACAG-----AGACTG	71
Qy	98	AlaAspGlyValGluTyrGlnPheThrPrrPheGlyValGlnValProGlyGlnIleIle	117
Db	72	TGTTTCG-----AGCAAGCCATCGTACCCGTGAACGGAGATGTTCCCGGACCCATCTTC	125
Qy	118	ArgValAlaArgGluGlyAspThrIleGluValAlaGlnPheSerAsnHisIleProAspSerIuYme	137
Db	126	TATGCTAGGGAGAGATGACACCCGTCCTCGTAGGGGTCTTAAC-----CGTGC	173
Qy	138	ProHisAsnValAspPheHis-----AlaAlaThr	147
Db	174	AAATACAAATGTCAACCATTCATGGCATGTGATCCGACGTGGAGACGAGGATGGCCGAC	233
Qy	148	GlyProGlyValGlyValAlaGluAlaSerPheThrAlaProGlyHisIleThrSerThrPheSer	167

Db 234 GGGCAGCAGTACATTTACCCATGCCCCGATC---CAACCGGGCCAAAGCATGTGTCACAT 290

Qy 168 PhelyalaLeu---GlnProGlyLeuTyTyrValTyThiSCysAlaValAlaproValGly 186

Db 291 TTCAACCATTCACGGGGCCAAACGGGCAACCCCTTCCTGTGGCATGCACACATA----- 338

Qy 187 MechistialaalaanglyMetTyrglyLeuIleLeuValGlnProGlyGlyLeu--- 205

Db 339 CTGTGGCTCAGGGCAACCCCTGCACGGAGCCATGTCACTTGGCCAAAGCTGGTGTCCCA 398

Qy 206 -----ProlySValAspLySgluTyTyTyrValMetGlnLyAspPheTyTThrLyS 222

Db 399 TACCCCTTCCTTAACCCCAAGAAAGTTGTTGTGTGTATGGCGGAATGGTAATGTC 458

Qy 223 GlyLyTyrgly-----GlnGlnGlyLeuGlnProPheAspMetGlu 236

Db 459 GATCAGACAGAGTGTGATCACTCAAGCCATCAAGTCCGGATTAGACACCG----- 506

Qy 237 LysAlaIleArgLysAspAlaGluTyTyrValValPheAsnGlySerValGlyAlaLeuThr 256

Db 507 -----AATGCTCTCGATGCTCACACAGATCAATGACCATCCAGGCCAAAGTCC 554

Qy 257 -----GlyGluAsnAlaLeuLySAlaLySValGlyGluThrValArgLeu 271

Db 555 AATTGCCCTTCCAGGGTGGATTTACGTTCCCTGTTGAGAGTGCAGAAAGTACATGCTG 614

Qy 272 PheValGlyAsnGlyTyProAsnLeuThrSerSerPheHisValIleGlyIuIlePhe 291

Db 615 CGAATCATCAACGCTGGCGCTCAATAGAGAGCTTCTTCAAGATTGCCGGGCAC----- 668

Qy 292 AspLysValHisPheGluGlyGlyGlyGlyGluAsnHisAsnIleGlnThrThrLeuIle 311

Db 669 -----CAGCTGACCATGTCG 683

Qy 312 ProAlaGlyGlyAlaIleAlaIleThrGluPheLySValAspValProGlyAspTyTyrValLeu 331

Db 684 GAGGTGCAGCCCACTCATCTCATCCAGCCCTTCAACACGACACG----- 725

Qy 332 ValAspHisAlaIlePheArgAlaPheAsnLySAlaLeuGlyIleLeuLySValGlu 351

Db 726 -----ATCGATGTTCACCT 740

Qy 352 GlyGluGluAsnHisGluIleTyTyrSerHisLySAlaGlnThrAspAlaValTyTyrLeuProGlu 371

Db 741 GGGCAAAACCAACATGCCCTCATCTTCACCGACACAGAGCTTGGCAAGTACATGTCGCC 800

Qy 372 Gly-----AlaProGlnAlaIleAspThrGlnGluAlaProLySThr 385

Db 801 GCCCTCCCTTTTATGACATCCCGATGCCGCTGCAGACATGACCGGACCGGCACA 857

RESULT 7

US-09-938-842A-1910

Sequence 1910, Application US/09938842A

Patent No. US20020160378A1

GENERAL INFORMATION:

APPLICANT: Harper, Jeff

APPLICANT: Krepes, Joel

APPLICANT: Wang, Xun

APPLICANT: Zhu, Tong

TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING

TITLE OF INVENTION: SAME, AND METHODS OF USE

FILE REFERENCE: SCRIPT300-3

CURRENT FILING DATE: 2001-08-24

PRIOR APPLICATION NUMBER: US 60/227,866

PRIOR FILING DATE: 2000-08-24

PRIOR APPLICATION NUMBER: US 60/264,647

PRIOR FILING DATE: 2001-01-16

PRIOR APPLICATION NUMBER: US 60/300,111

PRIOR FILING DATE: 2001-06-22

NUMBER OF SEQ ID NOS: 5379

SEQ ID NO 1910

LENGTH: 1713



```
Qy 161 GlyHisThrSerThrPheSerPheValalaLeu---GlnProGlyLeuTyValTyHis 179
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 506 GGGAGAGATTACGTACACTACACCTCGTGGGGGACGGCGGACGCTGTGGTGAC 565
Qy 180 CysAlaValAlaProValGlyMetHisLeaAlaAnglyMetTyGlyLeuIleVal 199
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 566 GCGCACATCTCC-----TGGCTGGCGCCCAACCGCTACGGGCGCCCTGTCATC 613
Qy 200 GlnProGlyGlyLeu-----ProGlyValaLeuGlyGlyTyValMet 215
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 614 CTGCCCAAGCTCGGGCTCCCTACCCGCTTCCCGGGCCCTACAGAGAGTCCCGTCATC 673
Qy 216 GlnGlyAerPheTyThrGlyGlyTyGlyGlyGlyGlyGlyGlyGlyGlyGly 235
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 674 TTCGTAGCTGTGCTGGCG-----GACACG 700
Qy 236 GlnValaAlaIleArgGlu-----AspAlaGlu 244
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 701 GAGGTGTGTATCAAGCAGCGCTTACGCTCGCGCTGGCCCAATGTCTGTAGCGCCAC 760
Qy 245 TyValValPheAnglySerValGlyAlaLeuThrGlyGlyAla----- 260
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 761 ACCATC-----AACGGCTGCCATAGCGGCTTACAACTGCTGCGCAAGACAGTAC 814
Qy 261 ---LeuValaValaValaGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 279
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 815 AAGCTGAAGTGAAGCCCGGGAAGACGATGCTGCTGCTGCTGCTGCTGCTGCTGCT 874
Qy 280 LeuThrSerPheHisValaIleGlyGlyIlePheAerGlyValHisPheGlyGly 299
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 875 GACAGAGCTCTTCTCTCGCTCGCCCAACAGCTGCTGCTGCTGCTGCTGCTGCTGCT 934
Qy 300 LyseGlyGlyAlaHisLeuIleGlnThrThrLeuIleProAlaGlyGlyAla-----Ala 317
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 935 TACGTCAAGCCTTTCACCGCTGACACGCTGCTCATCGCGCGGCGGCAACACCAAGCTG 994
Qy 318 IleThrGlnPheValaValaValaProGlyAerTyValaLeuValaPheAlaIlePhe 337
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 995 CTGCTCGCGCGGCAAGCCTCTTACCGCGGCGGCAAGTATGATCTCGCGCGGCGCTAC 1054
Qy 338 ArgAlaPheAerGlyGlyAlaLeuGlyIleLeuValaGlyGlyGlyGlyGlyGly 357
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1055 TCCACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1114
Qy 358 IleTySerHisValaGlnThrAerAlaValaTyLeuProGlyGlyAlaProGlyAlaIle 377
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1115 CTGTAC----- 1120
Qy 378 AspThrGlnGlnAlaProGlyThrProAlaProAlaLeuGlnGlnIleGlyAla 397
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1121 -----CCCGAGCGCGCGCGCGCGCTCCGC-CTCCGCGGAGACTTCACAGA 1164
Qy 398 GlyValaAlaThrTyAerSerAerGlyAlaAerHis----- 410
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1165 GGCCTGCGCGCTTACAGACGACCGCTCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1224
Qy 411 -----GlnProAerGlyGlyValaProAerAlaPheProPhe 424
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1225 CTTCACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1284
Qy 424 ValaAerSerAerTyLeuAerAla-----AspHisAlaAlaAlaSerIle 440
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1285 GGTGAGACAGGGGTTCTTCTCGCGGTGGGCGTGGGCGGCGGCGGCGGCGGCGGCGG 1344
Qy 440 ValaAlaAnglyLeuSerGly-----LysIleThrValaAnglyGlnTyGlyGly 458
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1345 CACGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1404
Qy 458 rValMetProAlaIleAlaLeu 465
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1405 CGTGCTCCCAACCAAGGGCTG 1426
```

```
US-10-115-563-17
: Sequence 17, Application US/10115563
: Publication No. US20030008307A1
: GENERAL INFORMATION:
: APPLICANT: Griffin, John H
: Greengard, Judith S
: TITLE OF INVENTION: METHODS FOR DIAGNOSING ACTIVATED PROTEIN
: C RESISTANCE ASSOCIATED WITH A FACTOR V GENETIC MUTATION
: AND COMPOSITIONS THEREOF
: NUMBER OF SEQUENCES: 28
: CORRESPONDENCE ADDRESS:
: ADDRESSER: The Scripps Research Institute, Office of
: Patent Counsel
: STREET: 10666 No. US20030008307A1h Torrey Pines Road, TPC 8
: CITY: La Jolla
: STATE: CA
: COUNTRY: USA
: ZIP: 92037
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: Patent Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/10/115,563
: FILING DATE: 02-Apr-2002
: CLASSIFICATION: <Unknown>
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/08/410,488
: FILING DATE: 24-MAR-1995
: ATTORNEY/AGENT INFORMATION:
: NAME: Fitting, Thomas
: REGISTRATION NUMBER: 34,163
: REFERENCE/DOCKET NUMBER: 449.0
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 619-554-6312
: TELEFAX: 619-554-2937
: INFORMATION FOR SEQ ID NO: 17:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2297 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: FEATURE:
: NAME/KEY: misc_difference
: LOCATION: replace(1614,"")
: OTHER INFORMATION: /label=N
: /note="Wherein 'N' is a guanine in a Factor V
: normal allele and an adenine in a Factor V mutant
: allele"
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: 1614
: OTHER INFORMATION: /note="Nucleotide position 1614
: below corresponds to nucleotide position 1691 of
: SEQ ID NO 13."
: SEQUENCE DESCRIPTION: SEQ ID NO: 17:
US-10-115-563-17

Alignment Scores:
Pred. No.: 0.00647 Length: 2297
Score: 128.50 Matches: 85
Percent Similarity: 33.02% Conservative: 55
Best Local Similarity: 20.05% Mismatches: 145
Query Match: 4.89% Indels: 139
DB: 14 Gaps: 21

US-10-088-045-2 (1-502) x US-10-115-563-17 (1-2297)
Qy 63 AspAlaIleValaThrHisAlaProGlyValaProProProValaParg----- 78
Db 1085 GAAGTCAATTGGAGCATGTGACCTGTGATATACAGCGCATATGACAAATAATACAGGCTCT 1144
```

```

QY 79 -----AspHisProAlaValValValLys 87
Db 1145 CAGCATTGGATTAATTTCTCAACCAATGGAAGAAACATTAAGAAAGTTATGATACACA 1204
QY 88 Met-----GluThrValGluValMetArgLeuAlaAsp 99
Db 1205 CAGTACGAGATGAGTCTTCCACCAACATACAGTATCCCAATATGAAA---GAAGAT 1261
QY 100 GlyValGluTyrGlnPheTyrPheGlyGlyGlnValProGlyGlnMetLeuVal 119
Db 1262 GGG-----ATTGGGTCTCTATATCAGAGCC 1288
QY 120 ArgGluGlyAspThrIleGluValGlnPheSerAsn-----HisProAspSerLys 136
Db 1289 CAGGTGAGAGACACACTCAAAATCGTGTCAAAAATATGGCCAGCCGCCCTATAGACTT 1348
QY 137 MetProHisAsnValAspPheHisAlaAlaThrGlyProGlyGlyGlyAlaGluAlaSer 156
Db 1349 TACCTTCATGAGTGAACCTTCTGCGCTTATGAAGATGA-----GTCAACTCTTCT 1399
QY 157 PheThrAlaProGlyHisThrSerThrPheSerPheLysAlaLeuGlnProGly---Leu 175
Db 1400 TTCACTCA---GGCAGGAACAACACCATG---ATCAGACAGATTCAACAGGGGAAACC 1453
QY 176 TyrValTyr-----HisCys 180
Db 1454 TATACTTATATGAGGAACATCTTAGAGTTGATGAACCCACAGAAATGATGCCCACTGC 1513
QY 181 AlaValAlaPro-----ValGlyMetHisIleAlaAsnGlyMetTyr 194
Db 1514 TTAAACAGACCACTACACTGACGTGACGTGACAGACATCGCTCTGGGCTATATA 1573
QY 195 GlyLeuIleLeuValGluProLysGlyLeuProLys-----Val 208
Db 1574 GGCCTACTTCTATCTGTAGACAGACAGATCCCTGACAGGCGNAGCATACAGAGGACGA 1633
QY 209 AspLysGluTyrTyrValMetGlnGlyAspPhe---TyrThrLysGlyLysTyrGlyGlu 227
Db 1634 GACATGGAACAGACAGCGGTGCTTGGCTGTGTTGATGAGAAACAAAGCTGTACCTTGAG 1693
QY 228 GlnGlyLeuGlnProPhe-----AspMetGlnLysAlaIleArgGluAspAlaGluTyr 245
Db 1694 GACAAATCAACAAGATTTGTGAAATCTCTGATGAGTGAACGTGATGACCCCAAGTTT 1753
QY 246 ValValPheAsnGlySerValGlyAlaLeuThrGlyGluAsnAlaLeuValVal 255
Db 1754 -----TATGATCAACCAATC 1768
QY 266 GlyGluThrValArgLeuPheValGlyAsnGlyGlyProAsnLeuThrSerPheHis 285
Db 1769 ATGAGCACTATCAATGCTATGTC-----CCTGAGACATTAATCT 1807
QY 286 ValIleGlyGluLeuPheAsp----- 292
Db 1808 ACTCTTGATTTGCTTGTATGACACTGTCAGTGCAGTCTTGTAGTGGGAGCCAG 1867
QY 293 -----LysValHisPheGluGlyGlyLys-----GlyGluAsnHisAsn 305
Db 1868 AATGAATTTTGGCCATCCATCCATCCATCCATCCATCCATCCATCCATCCATCCATCC 1927
QY 306 IleGlnThrThrLeuIleProAlaGlyAlaAlaIleThrGluPheLysValAspVal 325
Db 1928 GACACCTTGACCTTCTTCCCATGCGTGGAGATCTGTGACG----- 1969
QY 326 ProGlyAspTyrValLeuValAspHisAlaIlePheArgAlaPheAsnLysGlyAlaLeu 345
Db 1970 -----GTCAATGATGAT-----AATGTTGAACCTTG 1996
QY 346 GlyIleLeuLysValGluGlyGluAsnHisGluIleTyrSerHisLysGlnThrAsp 355
Db 1997 ATGTAACTTCCATGAATTCATGATCCAAAGAGAAAGCTGAGGCTGAATTCAGGAT 2056

```

```

QY 366 AlaValTyrLeuProGluGlyAlaProGlnAlaIleAspThrGlnGluAlaProLysTyr 385
Db 2057 GTTAAATGTAATCCAGATGATGATGAAGACTCATATGAGATTTTGAACCTCCAGATCT 2116
QY 386 ProAlaProAla-----AsnLeuGlnGlnIleLeuLysAlaGlyLys 399
Db 2117 ACAGTCATGGCTACACGGAATGATGATCGTTAGAACCTGGAAGATGAAGAGACTGAT 2176
QY 400 AlaThrTyrAsp 403
Db 2177 GCTGACTATGAT 2188

RESULT 10
US-10-115-563-27
; Sequence 27, Application US/10115563
; Publication No. US20030008307A1
; GENERAL INFORMATION:
; APPLICANT: Griffin, John H
; GREENGARD, Judith S
; TITLE OF INVENTION: METHODS FOR DIAGNOSING ACTIVATED PROTEIN
; C RESISTANCE ASSOCIATED WITH A FACTOR V GENETIC MUTATION
; AND COMPOSITIONS THEREOF
; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: The Scripps Research Institute, Office of
; Patent Counsel
; STREET: 10666 Nc. US20030008307A1th Torrey Pines Road, TPC 8
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/115,563
; FILING DATE: 02-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/410,488
; FILING DATE: 24-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Fitting, Thomas
; REGISTRATION NUMBER: 34,163
; REFERENCE/DOCKET NUMBER: 449.0
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-554-2937
; TELEFAX: 619-554-6312
; INFORMATION FOR SEQ ID NO: 27:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2297 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1614
; OTHER INFORMATION: /note="Nucleotide position 1614
; below corresponds to nucleotide position 1691 of
; SEQ ID NO 13"
; SEQUENCE DESCRIPTION: SEQ ID NO: 27:
US-10-115-563-27

Alignment Scores:
Pred. No.: 0.00647 Length: 2297
Score: 128.50 Matches: 85
Percent Similarity: 33.028 Conservative: 55
Best Local Similarity: 20.058 Mismatches: 145
Query Match: 4.89% Indels: 139
DB: 14 Gaps: 21

```





Pred. No.: 0.000647 Length: 2297  
 Score: 128.50 Matches: 85  
 Percent Similarity: 33.02% Conservative: 55  
 Best Local Similarity: 20.05% Mismatches: 145  
 Query Match: 4.89% Indels: 139  
 DB: 14 Gaps: 21

US-10-088-045-2 (1-502) x US-10-115-563-28 (1-2297)

QY Aspa1a1leValThrh1sa1aPProgluValPProProValaAspArg----- 78  
 Db 1085 GAAGTCATTGGAGCTATGCACTGTATATACCAAGATATGCAAAATAACAGCTCT 1144  
 QY 79 -----Asph1sProalalyValVallys 87  
 Db 1145 CAGCATTGGATAATTCTCAACCAATTGGAATACTTATAGAAAGTATATACACA 1204  
 QY 88 Met-----GluThyValGluValMetArgLeuAlaAsp 99  
 Db 1205 CAGTACGAAGATGATCTTCACCAACATACAGTATCCCAATATGAAA---GAAGAT 1261  
 QY 100 GlyValGluValGlnPheTrpThrPheGlyGlyGlnValProGlyGlnMetIleArgVal 119  
 Db 1262 GGG-----ATTGGGTCTCTATATACAGACC 1288  
 QY 120 ArgGluGlyAspThrIleGluValGlnPheSerAsn-----HisProAspSerLys 136  
 Db 1289 CAGGTCAAGACACACATCAAAATCGTGTAAAATATGCGCACGCCCTATATGACATT 1348  
 QY 137 MetProHisAsnValaPhePheHisAlaIleThrGlyProGlyGlyAlaGluAlaSer 156  
 Db 1349 TACCCTCATGAGTGCCTCTCGCTATAGAAATGAA-----GTCAACTCTTCT 1399  
 QY 157 PheThrAlaProGlyHisThrSerThrPheSerPheLysAlaLeuGlnProGly---Leu 175  
 Db 1400 TTCACTCTA---GCCAGAAACACACCATG---ATCAGAGCATGTCAACCGGGAAC 1453  
 QY 176 TyrValTyr-----HisCys 180  
 Db 1454 TATACTATAAGTGAACATCTTAGAGTTGATGACCAACAGAAATATGATGCCCAATGC 1513  
 QY 181 AlaValAlaPro-----ValGlyMetHisIleAlaAsnGlyMetTyr 194  
 Db 1514 TTAACAGACCATCTACAGTACGTGACATCATGAGACATCGCTCGGCTATA 1573  
 QY 195 GlyLeuIleLeuValGluProLysGluGlyLeuProLys-----Val 208  
 Db 1574 GGACTACTTCTAATCTGTAAGACGATCCCTGACACGCAAGGAATACAGAGGACGA 1633  
 QY 209 AspLysGluTyrTyrValMetGlnGlyAspPhe---TyrThrLysGlyLysTyrGlyGlu 227  
 Db 1634 GACATGCAACAGCAGCGTGTGCTGCTGTTTATAGAAACAAACCTGATCCTTGAG 1693  
 QY 228 GlnGlyLeuGlnProPhe-----AspMetGluLysAlaIleArgGluAspAlaGluTyr 245  
 Db 1694 GACAACATCAACAAGTTTGTGAATAATCTGATGAGTGAACGTATGACCCCAAGTTT 1753  
 QY 246 ValValPheAsnGlySerValGlyAlaLeuThrGlyLeuAlaLeuValaLysVal 265  
 Db 1754 -----TATGATATCAACATC 1768  
 QY 266 GlyGluThrValArgLeuPheValGlyAsnGlyGlyProAsnLeuThrSerPheHis 285  
 Db 1769 ATAGACATATCAATGCTATGTC-----CCTGAGACATAACT 1807  
 QY 286 ValIleGlyGluIlePheAsp----- 292  
 Db 1808 ACTCTTGATTCCTGCTTGATGACACTGCAGTGCACCTTCTAGTGGGAGCCAG 1867  
 QY 293 -----LysValHisPheGluGlyGlyLys-----GlyGluAsnHisAsn 305  
 Db 1868 AATGAATTTTGAACATTCATCTGAGCATCATATCTATGAAAGAGCATGAG 1927

QY 306 IleGlnThrThrLeuIleProAlaGlyGlyAlaAlaIleThrGluPheLysValaAspVal 325  
 Db 1928 GACACCTTGACCTCTTCCCATCGCGTGAAGATCTGTGACG----- 1969  
 QY 326 ProGlyAspTyrValLeuValaPheHisAlaIlePheArgAlaPheAsnLysGlyAlaLeu 345  
 Db 1970 -----GTCACAAATGAT-----AATGTGAAACTTGG 1996  
 QY 346 GlyIleLeuLysValaGluGlyGluGluAsnIleGlyIleTyrSerHisLysGlnThrAsp 365  
 Db 1997 ATGTTAATCTTCATGAATTTCTATCCAAAGAACAAAGCTGAGCTGAATTCAGGAT 2056  
 QY 366 AlaValTyrLeuProGluGlyAlaProGlnAlaIleAspThrGlnGluAlaProLysThr 385  
 Db 2057 GTTAAATGTATCCCATGATGATGAAAGACTCATATGAACTTTTGAACCTCCAGAAATCT 2116  
 QY 386 ProAlaProAla-----AsnLeuGlnGluGlnIleLysAlaGlyLys 399  
 Db 2117 ACAGTCATGGCTACAGGAAATGATGATCGTTTGAACCTGAAGATGAAGAGTGAAT 2176  
 QY 400 AlaThrTyrAsp 403  
 Db 2177 GCTGACTATGAT 2188  
 RESULT 12  
 US-09-880-107-2275  
 ; Sequence 2275, Application US/09880107  
 ; Patent No. US20020142981A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Horne, Darci T.  
 ; APPLICANT: Vockley, Joseph G.  
 ; APPLICANT: Scherf, Uwe  
 ; APPLICANT: Gene Logic, Inc.  
 ; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer  
 ; FILE REFERENCE: 44921-5028-WO  
 ; CURRENT APPLICATION NUMBER: US/09/880.107  
 ; CURRENT FILING DATE: 2001-06-14  
 ; PRIOR APPLICATION NUMBER: US 60/211,379  
 ; PRIOR FILING DATE: 2000-06-14  
 ; PRIOR APPLICATION NUMBER: US 60/237,054  
 ; PRIOR FILING DATE: 2000-10-02  
 ; NUMBER OF SEQ ID NOS: 3950  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 2275  
 ; LENGTH: 6909  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; OTHER INFORMATION: Genbank Accession No. US20020142981A1 M16967  
 US-09-880-107-2275  
 Alignment Scores:  
 Pred. No.: 0.00315 Length: 6909  
 Score: 128.50 Matches: 85  
 Percent Similarity: 33.02% Conservative: 55  
 Best Local Similarity: 20.05% Mismatches: 145  
 Query Match: 4.89% Indels: 139  
 DB: 10 Gaps: 21  
 US-10-088-045-2 (1-502) x US-09-880-107-2275 (1-6909)  
 QY 63 Aspa1a1leValThrh1sa1aPProgluValPProProValaAspArg----- 78  
 Db 1162 GAAGTCATTGGAGCTATGCACTGTATATACCAAGATATGCAAAATAACAGCTCT 1221  
 QY 79 -----Asph1sProalalyValVallys 87  
 Db 1222 CAGCATTGGATAATTCTCAACCAATTGGAATACTTATAGAAAGTATATACACA 1281  
 QY 88 Met-----GluThyValGluValMetArgLeuAlaAsp 99  
 Db 1282 CAGTACGAAGATGATCTTCACCAACATACAGTATCCCAATATGAAA---GAAGAT 1338

```

QY 100 GlyValGluTyrGlnPheThrPheGlyGluValProGlyGlnMetIleArgVal 119
Db 1339 GGG-----ATTGGGTCTTATATACAGGCC 1365
QY 120 ArgGluGlyAspThrIleGluValGlnPheSerAsn-----HisProAspSerLys 136
Db 1366 CAGGTGACGAGACACACTCAAAATCGTTCACAAAATATAGCCAGCCGCCCTATAGCATT 1425
QY 137 MetProHisAsnValAspPheHisAlaAlaThrGlyProGlyGlyValAlaGluAlaSer 156
Db 1426 TACCTTCATGAGTGCCTTCCTCGCTTATGAAGATGA-----GTCAACTCTTCT 1476
QY 157 PheThrAlaProGlyHisThrSerThrPheSerPheValAlaGlnProGly---Leu 175
Db 1477 TTCACCTCA---GGCAGAAACAACACCATG---ATCAGACGCTTCACACAGGGAAACC 1530
QY 176 TyrValTyr-----HisCys 180
Db 1531 TATACTTATATAGTGAACATCTTAGAGTTTGATGAACCCACAGAAATGATGCCAGTGC 1590
QY 181 AlaValAlaPro-----ValGlyMetHisIleAlaAsnGlyMetTyr 194
Db 1591 TTAAACAAGACCATCTACTACAGTACCTGACATCATGAGACATCGCTCGGCTTAATA 1650
QY 195 GlyLeuIleLeuValGluProLysGlyLeuProLys-----Val 208
Db 1651 GGACTACTCTTATATCTGTAGAGCAGATCCCTGCAGACGGCAGGAATACAGAGGCACGA 1710
QY 209 AspLysGluTyrTyrValMetGlnGlyAspPhe---TyrThrLysGlyLysTyrGlyGlu 227
Db 1711 GACATCGAAGCAGCAGCGCTGTCTGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1770
QY 228 GlnGlyLeuGlnProPhe-----AspMetGlyLysAlaIleArgGluAspAlaGluTyr 245
Db 1771 GACACATCAACAACATTTTGTGAAATCTCTGATGGGTAACCGATGAGCCCAAGTTT 1830
QY 246 ValValPheAsnGlySerValGlyAlaLeuThrGlyGluAsnAlaLeuValAlaVal 265
Db 1831 -----TANGAATCAACATC 1845
QY 266 GlyIuThrValAlaArgLeuPheValGlyAsnGlyGlyProAsnLeuThrSerSerPheHis 285
Db 1846 ATGAGCACTATCAATGGCTATGTG-----CCTGAGAGCATTAAT 1884
QY 286 ValIleGlyGluIlePheAsp-----GlyGluAsnHisAsn 292
Db 1885 ACTCTGTGATTCGTCTTGTGATGACACTGTCCAGTGGCACTTCTGATGTGGGACCCAG 1944
QY 293 -----LysValHisPheGluGlyGlyLys-----GlyGluAsnHisAsn 305
Db 1945 AATGAATTTTGAACCATCTCACTGCGACATCATTCATCTATGGAAGAGCATGAG 2004
QY 306 IleGlnThrThrLeuIlePheProAlaGlyGlyAlaIleThrGlyLysValAspVal 325
Db 2005 GACACCTTGACCTCTTCCCATCGTGGAGAACTGTGAGC-----2046
QY 326 ProGlyAspTyrValLeuValAspHisAlaIlePheArgAlaPheAsnLysGlyAlaLeu 345
Db 2047 -----ATGTTGAACCTGG 2073
QY 346 GlyIleLeuLysValGlyGlyGluAsnHisGluIleTyrSerHisLysGlnThrAsp 365
Db 2074 ATGTAACTTCAGAAATTCATGTCACAAAGAAAGCTGAGCTGAATTCAGGAT 2133
QY 366 AlaValTyrLeuProGluGlyAlaProGluAlaIleAspThrGlnGluAlaProLysThr 385
Db 2134 GTTAAATGATCCCAATGATGATGAAGACTCATATGAGATTTTGAACCTCCAGAAATCT 2193
QY 386 ProAlaProAla-----AsnLeuGlnGluGlnIleLysAlaGlyLys 399
Db 2194 ACAGTGATGCTACACGAAATATCATGATCGTTTAAACCTGAAGATGAAGAGATGAT 2253
QY 400 AlaThrTyrAsp 403

```

```

Db 2254 GCTGACTATGAT 2265
RESULT 13
US-10-115-563-13
Sequence 13, Application US/10115563
Publication No. US20030008307A1
GENERAL INFORMATION:
APPLICANT: Griffin, John H
Greenard, Judith S
TITLE OF INVENTION: METHODS FOR DIAGNOSING ACTIVATED PROTEIN
AND COMPOSITIONS THEREOF
C RESISTANCE ASSOCIATED WITH A FACTOR V GENETIC MUTATION
NUMBER OF SEQUENCES: 28
CORRESPONDENCE ADDRESS:
ADDRESSEE: The Scripps Research Institute, Office of
Patent Counsel
STREET: 10666 No. US20030008307A1th Torrey Pines Road, TPC 8
CITY: La Jolla
STATE: CA
COUNTRY: USA
ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/115,563
FILING DATE: 02-Apr-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/410,488
FILING DATE: 24-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Fitting, Thomas
REGISTRATION NUMBER: 34,163
REFERENCE/DOCKET NUMBER: 449.0
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-554-2937
TELEFAX: 619-554-6312
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 6925 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 175..6765
OTHER INFORMATION: /product= "Factor V"
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 91..174
FEATURE:
NAME/KEY: CDS
LOCATION: 91..6765
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1..8
FEATURE:
NAME/KEY: misc_feature
LOCATION: 6918..6925
OTHER INFORMATION: /standard_name= "EcoRI linker
nucleotide sequence"
SEQUENCE DESCRIPTION: SEQ ID NO: 13:
US-10-115-563-13
Alignment Scores:
Pred. No.: 0.00316 Length: 6925
Score: 128.50 Matches: 85
Percent Similarity: 33.02% Conservative: 55

```



Best Local Similarity: 20.05%  
 Query Match: 4.89%  
 DB: 14  
 Mismatches: 145  
 Indels: 139  
 Gaps: 21

US-10-088-045-2 (1-502) x US-10-115-563-26 (1-6925)

QY	63	AspAlaIleValThrHisAlaProGluValProProValAspArg-----	78
DB	1162	GAAGTCAATTGGAGCTATGCACTGTATATCCAGCAATATGGACAAATAACAGGTCT	1221
QY	79	-----AspHisProAlaValValValys	87
DB	1222	CAGCATTGGATATTCTCAAAACAAATTGAAACATTATAGAAAGTTATGACACA	1281
QY	88	Met-----GluThrValGluValMetArgLeuAlaAsp	99
DB	1282	CAGTACGAAGATGAGTCTTCACCAACATACAGTGAATCCCAATATGAAA---GAGAT	1338
QY	100	GlyValGluTyrGlnPheThrPheGlyGlnValProGlyGlnMetIleArgVal	119
DB	1339	GGG-----ATTGGGCTCTATTATCAGAGCC	1365
QY	120	ArgGluGlyAspThrIleGluValGlnPheSerAsn-----HisProAspSerLys	136
DB	1366	CAGGTCAAGACACACTCAAAATGCTTTCAAAATATGCGCAGCGCCCTATAGCATT	1425
QY	137	MetProHisAsnValAspPheHisAlaIleThrGlyProGlyGlyValGluAlaSer	156
DB	1426	TACCTCATGAGTGAACCTTCGCGCTATGAAATGAA-----GTCAACTCTTCT	1476
QY	157	PheThrAlaProGlyHisThrSerThrPheSerPheLysAlaLeuGlnProGly---Leu	175
DB	1477	TTCACTCA---GGCAGAACACACCATG---ATCAGACAGTTCAACAGGGGAAACC	1530
QY	176	TyrValTyr-----HisCys	180
DB	1531	TATACTTATTAAGTGAACATCTTAGATTGATGAACCCACAGAAATGATGCCAGTGC	1590
QY	181	AlaValAlaPro-----ValGlyMetHisIleAlaAsnGlyMetTyr	194
DB	1591	TTAACAGACCATTACTACAGTGAACGTGACATGAGACATCGCCTCGGGCTAATA	1650
QY	195	GlyLeuIleLeuValGluProLysGluLysLeuProLys-----Val	208
DB	1651	GGACTACTTCTTAATCTGTAAAGACAGATCCCTGAGCAGGCAAGATACAGAGGCGACA	1710
QY	209	AspLysGluTyrTyrValMetGlnGlyAspPhe---TyrThrLysGlyLysTyrGlyGlu	227
DB	1711	GACATCGAACAGCAGCGCTGCTGTGCTGTGATGAGAACAAAGCTGTAACCTTGAG	1770
QY	228	GlnGlyLeuGlnProPhe-----AspMetGluLysAlaIleArgGluAspAlaGluTyr	245
DB	1771	GACACATCAACAAAGTTTGTGAAATCTGTAGAGTGAACGTGATGACCCCAAGTTT	1830
QY	246	ValValPheAsnGlySerValGlyAlaLeuThrGlyLusAlaLeuLysAlaLysVal	265
DB	1831	-----TATGATCAAAACATC	1845
QY	266	GlyGluThrValArgLeuPheValGlyAsnGlyGlyProAsnLeuThrSerSerPheHis	285
DB	1846	ATGAGCACTATCAATGGCTATGTG-----CCTGAGAGCATTAAT	1884
QY	286	ValIleGlyGluIlePheAsp-----	292

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: September 5, 2003, 10:29:59 / Search time 3065 Seconds  
(without alignments)  
3980.700 Million cell updates/sec

Title: US-10-088-045-2  
Perfect score: 2626  
Sequence: 1 MSKPTLITKTLICALSALML.....NKGGQLSADVAKAKTKPN 502

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODE=frame+ p2n model -DEV=xlh  
-O=/cgn2.1/USPFO/spool/US10088045/runat\_04092003\_083142\_3928/app\_query.fasta.1.647  
-DB=EST -QFMT=fastap -SUFFIX=1st -MINMATCH=0.1 -ICOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blomsu62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=PCT -THR MAX=100 -THR MIN=0 -ALIGN=40 -MODE=LOCAL  
-OUTFMT=pco -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US10088045 @CCN 1.1 2810 @runat\_04092003\_083142\_3928 -NCPU=6 -ICPU=3  
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOC  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estlin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hlc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_liv:\*  
19: em\_gss\_pin:\*  
20: em\_gss\_vit:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_tod:\*  
26: em\_gss\_pig:\*  
27: em\_gss\_vrl:\*  
28: gb\_gss1:\*

29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	624	23.8	554	28	BH380443
2	581.5	22.1	513	28	BH401929
3	542	20.6	461	28	BH395696
4	303	11.5	263	28	BH375921
5	264.5	10.1	796	28	BH404724
6	210	8.0	795	28	BH380839
7	186.5	7.1	653	12	B1180641
8	182	6.9	719	29	B2337383
9	170.5	6.5	353	28	BH380607
10	158	6.0	492	29	B2716485
11	157	6.0	754	10	BG646015
12	155.5	5.9	700	28	BH378995
13	154.5	5.9	814	29	B2574542
14	152.5	5.8	413	9	AL377723
15	146	5.6	612	14	CB882241
16	145.5	5.5	501	14	CA553053
17	145.5	5.5	513	14	CA554730
18	145.5	5.5	515	14	CA551823
19	145.5	5.5	515	14	CA552719
20	145.5	5.5	1145	29	B2559318
21	144.5	5.5	567	10	BF648894
22	144.5	5.5	596	29	CC016516
23	143.5	5.5	360	14	CA550887
24	143.5	5.5	607	12	BH318801
25	142.5	5.4	623	9	AV822649
26	142.5	5.4	691	14	CA483187
27	142	5.4	752	28	AY079863
28	141.5	5.4	502	14	CA554547
29	141.5	5.4	600	9	AV826062
30	141.5	5.4	1269	29	B2560121
31	141	5.4	770	29	B2535086
32	140.5	5.4	515	14	CA553237
33	139.5	5.3	614	9	AV826269
34	139.5	5.3	659	10	BG448142
35	139.5	5.3	690	10	BF644576
36	139	5.3	571	9	AV826844
37	138.5	5.3	489	14	CA553614
38	138	5.3	645	10	BF650165
39	137	5.2	737	13	BH635352
40	136.5	5.2	612	9	AV826343
41	136.5	5.2	942	29	CNS070G7
42	134.5	5.1	563	12	B226252
43	133.5	5.1	523	9	AM695071
44	133	5.1	620	10	BE920041
45	133	5.1	650	10	BG135584

## ALIGNMENTS

RESULT 1  
BH380443  
LOCUS BH380443  
DEFINITION AG-ND-168117.TR ND-TRM Anopheles gambiae genomic clone AG-ND-168117  
ACCESSION BH380443  
VERSION BH380443.1 GI:17326585  
KEYWORDS  
SOURCE GSS.  
ORGANISM Anopheles gambiae (African malaria mosquito)  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;  
Anopheles.



Query Match: 22.14% Indels: 1  
DB: 28 Gaps: 1  
US-10-088-045-2 (1-502) x BH401929 (1-513)

QY 117 ILEAGVALARGGLUGLYSPHTRILEGLUVALGlnPheSerAsnHisProAspSerLys 136  
DB 7 ATCAGGGCAAGAGTGGTGACGCAAAATTCATTGCAATGAAATGAAACAGTACA 66  
QY 137 MetProHisAsnValAspPheHisAlaThrGlyProGlyGlyValAlaGlnLaser 156  
DB 67 TTCGCCGCAATATATACCTTCACTGCTGTAATGGCCGTGGCGGTGACAGAGTACT 126  
QY 157 PheThrAlaProGlyHisThrSerThrPheSerPheLysAlaLeuGlnProGlyLeuTy 176  
DB 127 TTGTAGCTCCGGGAAGAGAGCGGTATTAACTTCAAGCCCTTAATCCAGGTCTGTAC 186  
QY 177 ValTYrHisCysAlaValAlaProValAlaGlyMetHisAlaAlaSerGlyMetTy 196  
DB 187 GTATACCACTGTGCACCGCTCCGGTGGAAATGCATATTCCTAATGGATGATGATTA 246  
QY 197 ILEuValAlaGluProLysGlyLeuProLysValAspLysGlyTyTYrValMetGln 216  
DB 247 ATTCTGATAGAACCGAGAGGTGATTCCTTAAGATGATTAAGAGATTCTATATCATGCG 306  
QY 217 GlyAspPheTYrThyLysGlyLysTYrGlyGlyGlnGlyLeuGlnProPheAspMetGlu 236  
DB 307 GGAGATTCTTATACCAAGGTAAATACGAGACAAAGGTCTTCAGGAATTTGATATGAT 366  
QY 237 LysAlaAlaLeuGlyLysAlaGlyTyTYrValAlaPheAsnGlySerValGlyAlaLeuThr 256  
DB 367 AAGCAATTTGACAGATCTCAATATATGCTTTTCAATGATTAACAGACACACTTCG 426  
QY 257 GlyuAlaAlaLeuLysAlaLysValAlaGlyLysValAlaGlyLeuPheVal 273  
DB 427 ---GAGATATGACTGAAGAAAGGTGGTGAATAATGTAAGATTCCTTGT 474

RESULT 3 BH395696 461 bp DNA linear GSS 11-DEC-2001  
LOCUS AG-ND-105E5.TF ND-TAM Anopheles gambiae genomic clone AG-ND-105E5,  
DEFINITION genomic survey sequence.  
ACCESSION BH395696  
VERSION BH395696.1 GI:17341837  
KEYWORDS GSS.  
SOURCE Anopheles gambiae (African malaria mosquito)  
ORGANISM Anopheles gambiae  
REFERENCE Anopheles gambiae (Arthropoda; Hexapoda; Insecta; Pterygota;  
Eukaryota; Metazoa; Arthropoda; Diptera; Nematocera; Culicoidae;  
Anopheles.  
1 (bases 1 to 461)  
Hong, Y.S., Hogan, J.R., Wang, X., Sarkar, A., Sim, C., Loftus, B.J., Ren  
and Collins, F.H.  
Construction of a BAC library and generation of BAC end  
sequence-tagged connectors for genome sequencing of the African  
malaria mosquito Anopheles gambiae  
Mol. Genet. Genomics 268 (6), 720-728 (2003)  
22542063  
JOURNAL MEDLINE  
PUBMED 12653398  
COMMENT Other GSSs: AG-ND-105E5.TR  
Contact: Brendan J Loftus  
Department of Bukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0208  
Fax: 301 838 3543  
Email: b.loftus@tigr.org

This clone is from an A. gambiae BAC library (ND-TAM) provided by  
F.H. Collins and sequenced by The Institute for Genomic Research  
(TIGR). The BAC library was generated from A. gambiae PBST strain  
DNA. All DNA was extracted from newly hatched first instar larvae  
to minimize the inclusion of DNA from microorganisms that inhabit

the gut. The DNA is derived from mixed sexes of larvae. The BAC  
library was constructed at Texas A&M University BAC Center  
University, College Station, Texas 77843-2123, USA using a HindIII  
partial digest.  
Seq primer: M13 For  
Classes: BAC ends.

FEATURES  
source  
1..461  
location/Qualifiers

/organism="Anopheles gambiae"  
/mol\_type="genomic DNA"  
/strain="PBST"  
/db\_xref="taxon:7165"  
/clone="AG-ND-105E5"  
/clone\_1ib="ND-TAM"  
/note="Vector: pECBAC1; Site\_1: HindIII"

BASE COUNT 148 a 71 c 116 g 126 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 1,136-46 Length: 461  
Score: 542.00 Matches: 107  
Percent Similarity: 79.74% Conservative: 15  
Best Local Similarity: 69.93% Mismatches: 30  
Query Match: 20.64% Indels: 3  
DB: 28 Gaps: 0

US-10-088-045-2 (1-502) x BH395696 (1-461)

QY 116 MetIleArgValArgGlyGlyAspThrILEGLUVALGlnPheSerAsnHisProAspSer 135  
DB 3 CTTTATAGCGGCAAGAGTGGTGACGCAAAATTCATTGCAATGAAATGAAACAGT 62  
QY 136 LysMetProHisAsnValAspPheHisAlaThrGlyProGlyGlyValAlaGlnLaser 155  
DB 63 ACATTCGCCGCAATATATACCTTCACTGCTGTAATGGCCGTGGCGGTGACAGAGT 122  
QY 156 SerPheThrAlaProGlyHisThrSerThrPheSerPheLysAlaLeuGlnProGlyLeu 175  
DB 123 ACTTTGTAGAGTCCGGGATTTT-GCGGTATTTAATCTTCAAGCCCTTAATCCAGGTCTG 181  
QY 176 TYrValTYrHisCysAlaValAlaProValAlaGlyMetHisAlaAlaSerGlyMetTy 195  
DB 182 TACGTATACCACTGTGCACCGCTCCGGTGGAAATGCATATTCCTAATGGATGATGGA 241  
QY 196 ILEuValAlaGluProLysGlyLeuProLysValAspLysGlyTYrTYrValMet 215  
DB 242 TTAATTCGATAGAACCGAGAGGTGATTCCTTAAGATGATTAAGAGATTCTATATCATG 301  
QY 216 GlnGlyAspPhe-TYrThrLysGlyLysTYrGlyGlyGlnGlyLeuGlnProPheAspMe 235  
DB 302 CAGGAGATTTCAATATACCAAGGTAAATACGAGACAAAGTCTTCAGGAATTTGATAT 361  
QY 235 tGluLysAlaIleArgGlyuAspAlaGlyTYrValAlaPheAsnGlySerValGlyAlaLe 255  
DB 362 GGAATTAAGCAATTTGAGAAATCCCTGAATATGTGTTTCAATGATGTAAGACGAGACT 421  
QY 255 tThrGlyLysAlaLeuLysAlaLysValAlaGlyGlu 267  
DB 422 TCTGGAGATATGAACTGAAGA-AAGTTGGTGA 457

RESULT 4 BH375921 263 bp DNA linear GSS 10-DEC-2001  
LOCUS AG-ND-119L15.TF ND-TAM Anopheles gambiae genomic clone AG-ND-119L15,  
DEFINITION genomic survey sequence.  
ACCESSION BH375921  
VERSION BH375921.1 GI:17322063  
KEYWORDS GSS.  
SOURCE Anopheles gambiae (African malaria mosquito)  
ORGANISM Anopheles gambiae  
REFERENCE Anopheles gambiae (Arthropoda; Hexapoda; Insecta; Pterygota;  
Eukaryota; Metazoa; Arthropoda; Diptera; Nematocera; Culicoidae;  
Anopheles.

REFERENCE 1 (bases 1 to 263)  
 AUTHORS Hong, Y.S., Hogan, J.R., Wang, X., Sarkar, A., Sim, C., Loftus, B.J., Ren, C., Huff, E.R., Carlile, J.L., Black, K., Zhang, H.-B., Gardner, M.J., and Collins, F.H.  
 TITLE Construction of a BAC library and generation of BAC end sequence-tagged connectors for genome sequencing of the African malaria mosquito *Anopheles gambiae*  
 JOURNAL Mol. Genet. Genomics 268 (6), 720-728 (2003)  
 MEDLINE 22542063  
 PUBMED 12655398  
 COMMENT Other GSSs: AG-ND-119L15.TF  
 Contact: Brendan J Loftus  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0208  
 Fax: 301 838 3543  
 Email: b.loftus@igr.org  
 This clone is from an *A. gambiae* BAC library (ND-TAM) provided by F.H. Collins and sequenced by The Institute for Genomic Research (TIGR). The BAC library was generated from *A. gambiae* PEST strain DNA. All DNA was extracted from newly hatched first instar larvae to minimize the inclusion of DNA from microorganisms that inhabit the gut. The DNA is derived from mixed sexes of larvae. The BAC library was constructed at Texas A&M University BAC Center University, College Station, Texas 77843-2123, USA using a HindIII partial digest.  
 Seq primer: M13 For  
 Class: BAC ends.

FEATURES  
 source 1..263  
 Location/Qualifiers  
 /organism="Anopheles gambiae"  
 /mol\_type="genomic DNA"  
 /strain="PEST"  
 /db\_xref="taxon:7165"  
 /clone="AG-ND-119L15"  
 /clone\_lib="ND-TAM"  
 /note="Vector: pECBAC1, site\_1: HindIII"

BASE COUNT 76 a 48 c 72 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 6.23e-22 Length: 263  
 Score: 303.00 Matches: 57  
 Percent Similarity: 75.29% Conservative: 7  
 Best Local Similarity: 67.06% Mismatches: 20  
 Query Match: 11.54% Indels: 1  
 Gaps: 0

US-10-088-045-2 (1-502) x BH375921 (1-263)

QY 117 ||leargvalargluclyaprrhrllegluvalgnpheserashisproaseryls 136  
 |||||  
 DB 11 ATCAGGGCAAGAGTGGGTGACGAATTCATTGAAGATTAATAAACAAGTACA 70  
 |||||  
 QY 137 MetProHisAsnValAspPheHisAlaAlaThrGlyProGlyGlyValaGluAsp 156  
 |||||  
 DB 71 TTCGCCCAATATATACCTTCATGCTGTAAATGGCCCTGGCGGTGCAGAACTACT 130  
 |||||  
 QY 157 PheThrAlaProGlyHisThrSerThrPheSerPheValAlaGluGlnProGlyLeuTyr 176  
 |||||  
 DB 131 TTGTAGCTCCGGGAAGAAGAGCGTATTAACTTCAAAGCCCTAAATCCAGCTGTAC 190  
 |||||  
 QY 177 ValTyrHisCysAlaValAlaProValGlyMetHisAlaAlaAsnGlyMetTyr-GlyLe 196  
 |||||  
 DB 191 GATATACCACTGTGCGACCGCTTCGCGGATGATCATATTGTAATGGAGTGTATGGATT 250  
 |||||  
 QY 196 wileuValgu 200  
 |||||  
 DB 251 AATTCGATAGAA 263

RESULT 5  
 BH404724

LOCUS BH404724 796 bp DNA linear GSS 11-DEC-2001  
 DEFINITION AG-ND-139F12.TF.1 ND-TAM *Anopheles gambiae* genomic clone  
 ACCESSION AG-ND-139F12, genomic survey sequence.  
 VERSION BH404724  
 KEYWORDS BH404724.1 GI:17350940  
 SOURCE GSS.  
 ORGANISM *Anopheles gambiae* (African malaria mosquito)  
*Anopheles gambiae*  
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; *Anopheles*.

REFERENCE 1 (bases 1 to 796)  
 AUTHORS Hong, Y.S., Hogan, J.R., Wang, X., Sarkar, A., Sim, C., Loftus, B.J., Ren, C., Huff, E.R., Carlile, J.L., Black, K., Zhang, H.-B., Gardner, M.J., and Collins, F.H.  
 TITLE Construction of a BAC library and generation of BAC end sequence-tagged connectors for genome sequencing of the African malaria mosquito *Anopheles gambiae*  
 JOURNAL Mol. Genet. Genomics 268 (6), 720-728 (2003)  
 MEDLINE 22542063  
 PUBMED 12655398  
 COMMENT Other GSSs: AG-ND-139F12.TF.1  
 Contact: Brendan J Loftus  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0208  
 Fax: 301 838 3543  
 Email: b.loftus@igr.org  
 This clone is from an *A. gambiae* BAC library (ND-TAM) provided by F.H. Collins and sequenced by The Institute for Genomic Research (TIGR). The BAC library was generated from *A. gambiae* PEST strain DNA. All DNA was extracted from newly hatched first instar larvae to minimize the inclusion of DNA from microorganisms that inhabit the gut. The DNA is derived from mixed sexes of larvae. The BAC library was constructed at Texas A&M University BAC Center University, College Station, Texas 77843-2123, USA using a HindIII partial digest.  
 Seq primer: M13 Rev  
 Class: BAC ends.

FEATURES  
 source 1..796  
 Location/Qualifiers  
 /organism="Anopheles gambiae"  
 /mol\_type="genomic DNA"  
 /strain="PEST"  
 /db\_xref="taxon:7165"  
 /clone="AG-ND-139F12"  
 /clone\_lib="ND-TAM"  
 /note="Vector: pECBAC1, site\_1: HindIII"

BASE COUNT 278 a 132 c 176 g 210 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 3.12e-17 Length: 796  
 Score: 264.50 Matches: 55  
 Percent Similarity: 52.86% Conservative: 19  
 Best Local Similarity: 39.29% Mismatches: 49  
 Query Match: 10.07% Indels: 17  
 Gaps: 3

US-10-088-045-2 (1-502) x BH404724 (1-796)

QY 11 leuileCysAlaLeuSerAlaLeuMet---leuSerGlyCysSerAsnGlnAlaAspLys 29  
 |||||  
 DB 425 CTTCGCTGTTTGTGCTGCAATTCCTTAATGCTGTGAACAAACGCTCTGAA 484  
 |||||  
 QY 30 AlaAlaGlnProLysSerSerThrValAspAlaAlaAlaLysThrAlaAsnAlaAspAsn 49  
 |||||  
 DB 485 GCTTCACAA-----TCCGGAAGTACTGAAGGATTAAGTTCAGGGAATATCAGAA----- 535  
 |||||  
 QY 50 AlaAlaSerGlnLuhisGlnGlyLeuProValAlaAspAlaLeuValThrHisAla 69  
 |||||  
 DB 536 -----GAAGCCAACTGACATCACCCT 556





/lab host="DH10B (phage-resistant)"  
 /clone.lib="LRH (lin rhodamine and Hoechst dye)"  
 /note="Vector: pZLI; Site 1: SalI site; Site 2: EagI site;  
 LRH library was constructed from cDNA of primary bone  
 marrow cells depleted of lineage-committed cells and  
 enriched for primitive cells by FACS sorting for cells  
 with low level staining with rhodamine123 and Hoechst  
 33342 dyes. cDNA from 5000 cells derived from 30 mice,  
 5000 cells were were directionally cloned into SalI-EagI  
 restriction site of the (Ziplox (Gibco BRL). The ligated  
 cDNA fragments were transformed into DH10B host cells.  
 The original library had an initial plating complexity of  
 1.44X10<sup>7</sup> clones."
   
 BASE COUNT 118 a 200 c 208 g 125 t 2 others  
 ORIGIN

## Alignment Scores:

Pred. No.: 3,63e-09 Length: 653  
 Score: 186.50 Matches: 53  
 Percent Similarity: 48.25% Conservative: 16  
 Best Local Similarity: 37.06% Mismatches: 59  
 Query Match: 7.10% Indels: 15  
 DB: 12 Gaps: 6

US-10-088-045-2 (1-502) x B1180641 (1-653)

QY 366 AAlaValTyrLeuProGluGlyAlaProGlnAlaIleAspThrGlnAlaProLysThr 385  
 DB 613 GCCCGCTACCTCAAGAGCTTGGCGCG-----GTGATCCGAGNAGCGGCTCAAGCC 560  
 QY 386 ProAlaProAlaAsnLeuGln-----GluGlnIleLeysAlaGlyAla 400  
 DB 559 TATGAGCCGAGGCGGAGCCCTGTGGAAGGATGATACGACGAGCCGGGGCGTGC 500  
 QY 401 ThTThAspSerSerGlyAlaAlaGlyGlnProAspGlyLysGlyValProAsnAla 420  
 DB 499 GTCTATATTCAGACACTGGCGGCTGACCGATCCGACGCTATACCGCGGTG 440  
 QY 421 PheProProLeuAlaAsnSerAspTyrLeu---AsnAlaAspHisAlaArgAlaAsp 439  
 DB 439 TTCGCGGCGTGGCGGATATCCGCTGTCGCAATCGCGGAGCCGACGCTTGATCCAT 380  
 QY 440 IleValAlaAsnGly-----LeuSerGlyLysIleThrValAsnGlyAsnGlnTyrGlu 457  
 DB 379 ACCGTCCTCAAGGCGGACCTTGGCGGAGCCACACCGACCTTCGAAAC----- 329  
 QY 458 SerValMetProAlaIleAla-----LeuSerAspGlnGlnIleAlaAsnValIleThr 475  
 DB 328 TTCACCATGCGCGCTTGGCTGGCGATGTCGACAGAGTGGCGGACGTGGCGGT 269  
 QY 476 TyrThrLeuAsnSerPheGlyAsnLysGlyGlnLeuSerAlaAspAspValAlaLys 495  
 DB 268 TTCATTCGCGGCGGAGTGGGGTATATCAGCGCGGACGATGATGCGGCG 209  
 QY 496 AlaLysLys 498  
 DB 208 TTGCGCAAG 200

RESULT 8 BZ337383 719 bp DNA linear GSS 06-NOV-2002  
 LOCUS BZ337383  
 DEFINITION bicolor genomic clone 1a87a06 5', genomic survey sequence.  
 ACCESSION BZ337383  
 VERSION BZ337383.1 GI:24731977  
 KEYWORDS GSS.  
 SOURCE Sorghum bicolor (sorghum)  
 ORGANISM Sorghum bicolor  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACAD  
 clade; Panicoideae; Andropogoneae; Sorghum.  
 1 (bases 1 to 719)  
 Rabinowitz,P.D., O'Shaughnessy,A.L., Balija,V., Dedhia,N.,

TITLE  
JOURNAL  
COMMENT

Katzenburger, F., King, L., Miller, B., Muller, S., Nascimento, L.,  
 Zlatavern, T., Palmer, L., McCombie, W.R. and Martienssen, R.A.  
 Genomic shotgun sequences from Sorghum bicolor (methyl-filtered)  
 Unpublished  
 Contact: W. Richard McCombie  
 Lita Annenberg Hazen Genome Sequencing Center  
 Cold Spring Harbor Laboratory  
 PO Box 100, Cold Spring Harbor, NY 11724, USA  
 Tel: 516 367 8884  
 Fax: 516 367 8874  
 Email: mcombie@cshl.org  
 Plate: 1a87 row: a column: 06  
 Seq primer: -21M3univrev  
 Class: Shotgun  
 High quality sequence stop: 719.  
 Location/Qualifiers

FEATURES  
source

1..719  
 /organism="Sorghum bicolor"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:4558"  
 /clone="1a87a06"  
 /lab host="JMI07 or DH5a"  
 /note="Site 1: Xba I; Site 2: Xba I; The vector was filled in  
 digested with Xba I and one nucleotide was added by fill in  
 in the recessive 3' end. The genomic DNA was nebulized,  
 end repaired, adaptor ligated and size fractionated using  
 sephadex. The resulting fragments were between 0.8 and 3  
 kb and were cloned into the vector (x/y reads in M13mp19,  
 b/g reads in pUC19). The same ligation was transformed in  
 either JMI07 or DH5a."
   
 BASE COUNT 148 a 232 c 214 g 125 t  
 ORIGIN

## Alignment Scores:

Pred. No.: 1.23e-08 Length: 719  
 Score: 182.00 Matches: 44  
 Percent Similarity: 55.14% Conservative: 15  
 Best Local Similarity: 41.12% Mismatches: 42  
 Query Match: 6.93% Indels: 6  
 DB: 29 Gaps: 4

US-10-088-045-2 (1-502) x BZ337383 (1-719)

QY 396 LysAlaGlyLysAlaThrTyrAspSerAsnGlyAlaGlyHisGlnProAspGlyLys 415  
 DB 42 AAGCGGCGCGCTGGGTATATGACACACTGGCGGCTTGCACCGACCGGCGCAC 101  
 QY 416 GlyValProAsnAlaPheProProLeuAlaAsnSerAspTyrLeu---AsnAlaAspHis 434  
 DB 102 GGTACACGCGCTGATATCCCGGCGCTGGCGGACCCCGCTTGCAGACGGCGGATGCG 161  
 QY 435 AlaArgAlaAspSerIleValAlaAsnGlyLeuSerGlyLysIleThrValAsnGlyAsn 454  
 DB 162 ACGTCTGATCATATATCTGTTGAACGCG-----GGACACTTGGCGGACCCACACG 215  
 QY 455 GlnTyrGluSer---ValMetProAlaIleAla-----LeuSerAspGlnGlnIleAla 471  
 DB 216 GCGCGCTACCTTACACCATGCGCGCTTGGCTGGCGGCTTGGCGGAGAGTGGCG 275  
 QY 472 AsnValIleThrTyrThrLeuAsnSerPheGlyAsnLysGlyGlnLeuSerAlaAsp 491  
 DB 276 GATGTGTACGTTTCGTCGCTGGCGGATGTTGGGGTATACAGGCGACCGCTGTAAGGCCAGT 335  
 QY 492 AspValAlaLysAlaLysLys 498  
 DB 336 GACGTGGCGGACCTGGCGCAAG 356

RESULT 9 BH380607 353 bp DNA linear GSS 10-DEC-2001  
 LOCUS BH380607/c  
 DEFINITION AG-ND-164017.TF ND-TMW Anopheles gambiae genomic clone AG-ND-164017  
 , genomic survey sequence.



QY 145 -----AlaAlaThrGlyProGlyGlyValAlaGluAlaSerPheThrAlaPro 160  
 DB 196 CCGGACGCCCGTGGGCGCCACGACGAGGAGGCGTACGACGAGCCCATC---CTCCCC 252  
 QY 161 GlyHisThrSerThrPheSerPheLeuValAlaGluProGlyLeuValTyrHisCys 180  
 DB 253 GGGGACACCTTCACTACGCTTCGTGTGAGACCGCCGCGGACGATACATGTCAC--- 309  
 QY 181 AlaValAlaProValGlyMethHisIleAlaSerGlyMethTyrGlyLeuValGlu 200  
 DB 310 -----GGCCACTACGACGATGACGCGCTCCGCGGAGCTCAACGCGCTCATGCTGTGGCC 363  
 QY 201 ProLysGluGlyLeuPro-----LysValAlaSerGlyTyrTyrValMet 215  
 DB 364 GGGGCGCCCGCGCGCGCCGACCGCGCGCTTCCTCCGCTACGACGCGACGACGCTGTG 423  
 QY 216 GlnGlyAspPheThrTyrGlyGlyTyr 225  
 DB 424 CTCACGACTGCTGTGCATTAAGACACCTAC 453  
 RESULT 11  
 BG646015 754 bp mRNA linear EST 24-APR-2001  
 LOCUS EST507634 KV3 Medicago truncatula cDNA clone pKV3-48A4 5' end, mRNA  
 DEFINITION sequence.  
 ACCESSION BG646015  
 VERSION BG646015.1 GI:13781127  
 KEYWORDS EST.  
 SOURCE Medicago truncatula (barrel medic)  
 ORGANISM Medicago truncatula  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids  
 ; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae;  
 Medicago.  
 1 (bases 1 to 754)  
 Vandenbosch,K., Endre,G., Hur,J., Beremand,P., Town,C.D., Bowman  
 C.L., Craven,M.B., Cho,J. and Fraser,C.M.  
 ESTs from roots of Medicago truncatula 72 h after Rhizobium  
 inoculation, 2001  
 JOURNAL Unpublished  
 COMMENT Contact: Vandenbosch K  
 Department of Plant Biology  
 University of Minnesota  
 220 Biosci Center, 1445 Gortner Ave, St. Paul, MN 55108, USA  
 Tel: 612 624 2755  
 Fax: 612 625 1738  
 Email: kvandenb@cds.umn.edu  
 M394422e TIGR sequence name: MTECU02TK More information is  
 available at: www.medicago.org  
 Seq primer: Skmod (CTA gaa cta gtc gat cc).  
 FEATURES  
 source  
 1..754  
 /organism="Medicago truncatula"  
 /mol\_type="mRNA"  
 /cultivar="genotype Al7"  
 /db\_xref="taxon:3880"  
 /clone="pKV3-48A4"  
 /tissue\_type="Seedling roots"  
 /dev\_stage="3 days post-inoculation with Sinorhizobium  
 meliloti"  
 /lab\_host="E. coli strain XLOLR"  
 /clone\_1ib="KV3"  
 /note="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2:  
 XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA  
 was directionally ligated into the Unizap XR vector from  
 Stratagene and packaged using Gigapack III Gold packaging  
 extracts. Plasmids containing cDNA inserts were excised  
 from the recombinant lambda-Zap phage using Ex-aseist  
 helper phage and propagated in XLOLR cells."

Pred. No.: 5,51e-06 Length: 754  
 Score: 157.00 Matches: 72  
 Percent Similarity: 38.70% Conservative: 29  
 Best Local Similarity: 27.59% Mismatches: 100  
 Query Match: 5.98% Indels: 60  
 DB: 10 Gaps: 14  
 US-10-088-045-2 (1-502) x BG646015 (1-754)  
 QY 99 AspGlyValGlyTyrGlnPheThrPheGlyGlyGlnValProGlyGlnMetIleArg 118  
 DB 5 GATTGTAAAGACATGTTGTGATGGAAATCAATGCCAATTTCCAGGCCCACTATTAGA 64  
 QY 119 ValArgGluGlyAspThrIleGluValGlnPheSerAsn-----HisProAspSerLys 136  
 DB 65 GCTGAAGTTGGTGAACACTCTTGTATTGACCTCACACACAGCTCATACAGAGGAACT 124  
 QY 137 MetProHisAsnValAspPheHisAlaAlaThrGlyProGly-----Gly 151  
 DB 125 GTTATTTCAC-----TGCGATGGAATCAGACAGTTTGGAACTCTGGCTGATGA 175  
 QY 152 GlyAlaGluAlaSerPheThrAla-----ProGlyHisThrSerThrPheSerPheLys 169  
 DB 176 ACTGCTGCATATTCATCAATGCTCTATTAATCCAGAGAAACTTTCAATACAAATTCAAA 235  
 QY 170 AlaLeuGlnProGlyLeuTyrValTyrHisCysAlaValAlaProValGlyMethHisIle 189  
 DB 236 GTTGACAGGCTGTGTACATATTCTATCT-----GGACACTATGTATGCAGAA 286  
 QY 190 AlaAsnGlyMetTyrGlyLeuIleLeuValGluProLysGluGlyLeuProLysValAsp 209  
 DB 287 GCAGACAGGTTGTATGTTCTCTTAATAGTGAT-----TTACCAAGAGCCAA 334  
 QY 210 LysGlyTyrTyrValMetGlnLysAspPhe-----TyrThr 221  
 DB 335 AAGAGGCCATTTCAATACATGATGATGATTCCTCTAGTATGATCTGGACACACA 394  
 QY 222 LysGlyLysTyrGlyGluGlnGlyLeuGlnProPheAspMetGluLysAlaIleArgGlu 241  
 DB 395 AGTTCAATCAACAAAGAGGTGGCTCTCTTCTGCCCAATG---AGATGATTTGGTGA 451  
 QY 242 AspAlaGluTyrValVal-----PheAsnGlySerValGlyAlaLeuThr 256  
 DB 452 CCACAGAGCTGTCTATCAATGACAGACAGACAGTTCAATTTGTTCCCTACATCAATAT 511  
 QY 257 GlyLysAsnAlaLeu-----LysAlaLysValGlyGluThrValArgLeuPheVal 273  
 DB 512 GGGAGCACCAACCTACCCCAATGCAATTTGAAAGGTGTGAAGATGT----- 559  
 QY 274 GlyAsnGlyGlyProAsnLeuThrSerSerPheHisValIleGlyGluIlePheAspLys 293  
 DB 560 -----GCACCCCAAGATT----- 571  
 QY 294 ValHisPheGluGlyGlyLysGlyLysAsnHisIle-----GlnThrThrLeuIlePro 312  
 DB 572 CTTCACGTGAGGACAAAAGACCTATAGATATAGATGTCTAGTACAGCTCTGGCT 631  
 QY 313 AlaGlyGlyAlaAlaIleThrGluPheLys-----ValAspValProGlyAspTyr 329  
 DB 632 TCTCTCACTGGCCATTCAATCAACAACTATTATTAGTGAAGACAGATGGAATCTAT 691  
 QY 330 Val 330  
 DB 692 GTG 694  
 RESULT 12  
 BH378995 700 bp DNA linear GSS 10-DEC-2001  
 LOCUS AG-ND-162A8.TF ND-TAM Anopheles gambiae genomic clone AG-ND-162A8,  
 DEFINITION genomic survey sequence.  
 ACCESSION BH378995  
 VERSION BH378995.1 GI:17325137  
 KEYWORDS GSS.

SOURCE  
ORGANISM Anopheles gambiae (African malaria mosquito)

REFERENCE  
AUTHORS 1 (bases 1 to 700)  
Hong, Y.S., Hogan, J.R., Wang, X., Sarkar, A., Sim, C., Loftus, B.J., Ren, C., Huff, E.R., Carille, J.L., Black, K., Zhang, H.-B., Gardner, M.J., and Collins, F.H.  
Construction of a BAC library and generation of BAC end sequence-tagged connectors for genome sequencing of the African malaria mosquito *Anopheles gambiae*  
Mol. Genet. Genomics 268 (6), 720-728 (2003)

JOURNAL  
MEDLINE 22542063  
PUBMED 12655398

COMMENT  
Other GSSs: AG-ND-162A8.TR  
Contact: Brendan J Loftus  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0208  
Fax: 301 838 3543  
Email: b.loftus@tigr.org

TITLE  
This clone is from an *A. gambiae* BAC library (ND-TAM) provided by F.H. Collins and sequenced by The Institute for Genomic Research (TIGR). The BAC library was generated from *A. gambiae* PEST strain DNA. All DNA was extracted from newly hatched first instar larvae to minimize the inclusion of DNA from microorganisms that inhabit the gut. The DNA is derived from mixed sexes of larvae. The BAC library was constructed at Texas A&M University BAC Center University, College Station, Texas 77843-2123, USA using a HindIII partial digest.  
Seq primer: M13 For  
Class: BAC ends.

FEATURES  
source  
1..700  
/organism="Anopheles gambiae"  
/mol\_type="genomic DNA"  
/strain="PEST"  
/db\_xref="taxon:7165"  
/clone="AG-ND-162A8"  
/clone\_lib="ND-TAM"  
/note="vector: pECBAC1; site\_1: HindIII"

BASE COUNT 185 a 156 c 116 g 243 t  
ORIGIN

Alignment Scores:  
Pred. No.: 7,15e-06 Length: 700  
Score: 155.50 Matches: 38  
Percent Similarity: 50.94% Conservative: 16  
Best Local Similarity: 35.85% Mismatches: 34  
Query Match: 5,92# Indels: 18  
DB: 28 Gaps: 3

US-10-088-045-2 (1-502) x BH378995 (1-700)

QY 11 LeuUcYsAlaLeuSerAlaLeuMet---LeuSerGlyCySerAsnGlnAlaAsp-Ly 29  
DB 212 CTTCGCTGTTGTTGTCGCAATTTTCACCTTGATGATGCTTGTAACAAACCTTCCTGA 213

QY 29 sAlaAlaGlnProLysSerSerThrValAspAlaAlaAlaLysThrAlaAsnAlaAspAs 49  
DB 212 AGCTTCACAA-----TCCGGAAGTACTGAGGTAATTACTGTCAGGGAATGCGAGA-- 161

QY 49 nAlaAlaSerGlnGluHisGlnGlyGluLeuProValIleAspAlaIleValThrHisAl 69  
DB 160 -----GAAGCCAAACTGACATCAC 141

QY 69 aProGluValProProValAspArgAspHisProAlaLysValValValLysMetG 89  
DB 140 TTCCAATGCGCTTGAACCAATAGGAACAGAGCTGCAAAAAGCGATGTGACGCTGGA 81

QY 89 uThrValGluLysValMetArgLeuAlaAspGlyValGluTyrGlnPheTrpThrPheG 109

DB 80 AACATATGAAAAAACAAGGTGAATGTGACAGACCGAACCAATATTAATCTTGACCTTTG 21  
QY 109 YGtGlnValProGly 114  
DB 20 CGGAACAGTCCGGGA 5

RESULT 13  
BZ574542/c  
LOCUS  
DEFINITION  
BZ574542 814 bp DNA linear GSS 17-DEC-2002  
mesh\_3719.y2 mesh *Pseudomonas aeruginosa* genomic clone mesh\_3719,  
genomic survey sequence.  
BZ574542  
ACCESSION  
BZ574542  
VERSION  
BZ574542.1 GI:27209603  
KEYWORDS  
GSS.  
SOURCE  
ORGANISM  
*Pseudomonas aeruginosa*  
*Pseudomonas aeruginosa*  
*Bacteria*; *Proteobacteria*; *Gammaproteobacteria*; *Pseudomonadales*;  
*Pseudomonadaceae*; *Pseudomonas*.  
1 (bases 1 to 814)  
Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M.,  
Burns, J.L., Kaul, R., and Olsen, M.V.  
Whole-genome-sequence variation among multiple isolates of  
*Pseudomonas aeruginosa* library  
J. Bacteriol. (2002) In press  
Contact: Chris K. Raymond  
Genome Center  
University of Washington  
Box 352145, Seattle, WA 98105-2145, USA  
Tel: 2062216954  
Fax: 2066857244  
Email: ckraymond@u.washington.edu  
Class: shotgun.

FEATURES  
source  
1..814  
/organism="Pseudomonas aeruginosa"  
/mol\_type="genomic DNA"  
/strain="MSH"  
/db\_xref="taxon:287"  
/clone="mesh\_3719"  
/clone\_lib="MSH"  
/note="Environmental isolate. Whole genomic shotgun library."

BASE COUNT 127 a 243 c 283 g 160 t 1 others  
ORIGIN

Alignment Scores:  
Pred. No.: 1,12e-05 Length: 814  
Score: 154.50 Matches: 53  
Percent Similarity: 42.31% Conservative: 35  
Best Local Similarity: 25.48% Mismatches: 79  
Query Match: 5,88# Indels: 41  
DB: 29 Gaps: 9

US-10-088-045-2 (1-502) x BZ574542 (1-814)

QY 325 ValProGlyAAspTyrValLeu-----ValAspHisAlaIlePheArgAlaPhe 340  
DB 774 GTTCCGGGCAAGTTTACTCGACGCGCAACCGTGAACAT-----TGGCAACCGCTG 721

QY 341 Asn-----LysGlyAlaLeuGlyIleLeuLysValGluGlyGlu 353  
DB 720 AGCTTGCCCACTTGTGAGCGGTGAGAACCCCGTCAGCTTCTGAAGACC---GGGCA 664

QY 354 GluAsnHisGluIleTyrSerHisLysGlnThrAspAlaValTyr----- 368  
DB 663 ANCCGCTTGCCGACCGTCCGCGAGCATGACCGATGCATCCACCAAGCAGCAGCAGC 604

QY 369 -----LeuProGluGlyAla 373  
DB 603 TTCAGCGACGAGCATCTGCTGCGCATCGCCAGCTAAGTCCCTCCGCGCGCAAG 544

QY 374 ProGlnAlaIleAspThrGlnGluAlaProLysThrProAlaProAlaAsnLeuGlnGlu 393

```

Db      543 GACGACCTGCCATGCCCGACGACGACGCCCACTGCGACACCGCTGCACCTGTACAGC 484
Oy      394 GlnlleLysalaGlyLysalaThrTyraSerSerAsnCyalaAlaCyshiSglnProAsp 413
Db      483 -----TCGCGGGGGCGGCTCGGCTACGCGCAAGTCTGCTCCGACCTGCACCGCAAGAC 430
Oy      414 GlyLysGlyValProAsnAlaPheProPheLeuAlaAsnSerAspTyreLeuAlaAsp 433
Db      429 GCGAGCGGGCGCTCGGTATGTTCCGCGCTGCGCGCAACCCACGCGCTTG--- 373
Oy      434 HisAlaArgAlaAlaSerIleValAlaAsnGlyLeuSerGlyLysIleThrValaAsnGly 453
Db      372 ---GCCAACCCGACGACGCTACCTCATACCTTACCGCTGCAAAACCGCGAGACC 316
Oy      454 AsnGlnTyreLysSerVal-----MetProAlaIleAla---LeuSerAspGlnGlnIle 470
Db      315 GCMAACCACTCGCGGCTCTACACCATGCGCGCTGCGCCAGCTGGAAAGACCGCAATC 256
Oy      471 AlaAsnValIleThrTyreThrLeuAsnSerPheGlyAsnLysGlyGlnLeuSerAla 490
Db      255 GCGGAGATCCTCGACTTCCTCCGACGACAGTTGGGCAACAGGGTTCGTGATCGATGCC 196
Oy      491 AspAspValAlaLysAlaLysAlaLys 498
Db      195 GCGCAGCTGAAGAACTCGCCGACG 172

RESULT 14
AL377723 413 bp mRNA linear EST 03-AUG-2000
LOCUS MCB833D10F1 MCB8 Medicago truncatula cDNA clone MCB833D10 T3, mRNA
DEFINITION
ACCESSION AL377723
VERSION AL377723.1 GI:9677475
KEYWORDS EST.
SOURCE Medicago truncatula (barrel medic)
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
; eucosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae;
Medicago.
1 (bases 1 to 413)
Journet, E.P., Crepeau, H., van Nulmen, D., Gouzy, J., Jallion, O.,
Niebel, A., Carreau, V., Chateigner, O., Kahn, D., Giamnazi, Pearson
, V. and Gamas, P.
Medicago truncatula ESTs from Sinorhizobium meliloti-induced root
nodules
Unpublished
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seque@genoscope.cns.fr, Web : www.genoscope.cns.fr
Contact : Pascal Gamas and Etienne-Pascal Journet, Laboratoire de
Biologie Molculaire des Relations Plantes-Microorganismes,
CNRS-INRA, BP 27 31326 Castanet-Tolosan Cedex, France (Email :
mc-set@oulouse.inra.fr Website :
http://sequence.toulouse.inra.fr/McTruncatula.html).
Location/Qualifiers
1. 413
/organism="Medicago truncatula"
/mol_type="mRNA"
/cultivar="Jemalong"
/db_xref="taxon:3880"
/clone="MCB833D10"
/tissue_type="symbiotic root nodules"
/dev_stage="harvested 4 days post inoculation with
Sinorhizobium meliloti"
/clone_lib="MCBB"
/clone_vector="pBluescript pSK, Site.1: EcoRI, Site.2:
XhoI; Plants were grown in an aeroponic chamber on
nitrogen-rich medium for 21 days. Three days before
inoculation with Sinorhizobium meliloti, the medium was
replaced by N-free medium. Root nodules (+ short adjacent

```

```

root segments) were harvested 4 days post inoculation.
cDNA was prepared from polyA+ enriched RNA. The cDNA was
directionally ligated into Uni-zap XR vector from
Stratagene and packaged using Gigapack Gold packaging
extracts. Plasmids containing cDNA inserts were
mass-excised from phage stocks using Exsist helper phage
and propagated in SOLR cells. Clone ordering and
sequencing was performed by the Centre National de
Sequencage (Genoscope, Evry, France)."

BASE COUNT      132 a      77 c      92 g      112 t
ORIGIN

Alignment Scores:
Pred. No.:      7,16e-06      Length:      413
Score:          152.50      Matches:      45
Percent Similarity: 45.65%      Conservative: 18
Best Local Similarity: 32.61%      Mismatches: 56
Query Match:      5.81%      Indels:      19
DB:              9      Gaps:      6

US-10-088-045-2 (1-502) x AL377723 (1-413)
Oy      91 ValGluLysValMetArgLeuAlaAspGlyValGluTyreGlnPheThrPheGlyGly 110
Db      25 GTGAATACATATACAAAAGCCAGATTTGTAAGAGCATGTTGTGATGGGAATCATGCG 84
Oy      111 GlnValProGlyGlnMetIleArgValArgGluGlyAspThrIleGluValGlnPheSer 130
Db      85 CAATTTCAGGCCCACTATTAGACTGGAAGTTGGTGAACACTCTGTTATACCTCAC 144
Oy      131 Asn-----HisProAspSerLysMetProHisAsnValAspPheHisAlaAlaThrGly 148
Db      145 AACAGCTCCATACAGAGGAACTGTTATTC-----TGCAATGAAATACAGACAG 195
Oy      149 ProGly-----GlyGlyAlaGluAlaSerPheThrAla-----ProGly 161
Db      196 TTGGAACTCCTTGGGCTGATGAACTGCTGCATATCTCATATGCTAATATCCAGCA 255
Oy      162 HisThrSerThrPheSerPheLysAlaLeuGlnProGlyLysTyreValTyreHisCysAla 181
Db      256 GAACCTTTCAATACAAATTCAAAGTTGACAGCGCTGGACATATTTCTATCAT----- 309
Oy      182 ValAlaProValGlyMetHisIleAlaAsnGlyMetTyreGlyLeuIleLeuValGluPro 201
Db      310 ---GCACACTAGCTGATGCAAGACGACGAGGTTGATGCTCTATATATGAT--- 363
Oy      202 LysGluGlyLeuProLysValAspLysGlyTyreValMetGlnGlyAspPhe 219
Db      364 -----TTACCAAGAGCCAAAGAGAGCCATTTCATTACGATGTCAGTTTC 408

RESULT 15
CB882241 612 bp mRNA linear EST 23-APR-2003
LOCUS HL01E24w HL Hordeum vulgare subsp. vulgare cDNA clone HL01E24
DEFINITION
ACCESSION CB882241
VERSION CB882241.1 GI:30084234
KEYWORDS EST.
SOURCE Hordeum vulgare subsp. vulgare
ORGANISM Hordeum vulgare subsp. vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidaeae
; Triticeae; Hordeum.
1 (bases 1 to 612)
Stein, N., Potokina, E. and Graner, A.
Barley ESTs from pooled RNAs of germinating seed (enriched for full
length transcripts)
Unpublished
Contact: Stein Nils
Molecular Markers Group, Department Genbank
Institute of Plant Genetics and Crop Plant Research (IPK)
Corrensstr. 3, 06466, Gatersleben, Germany
Tel: 039482-5522

```

Fax: 039482-5595  
Email: stein@pk-gatersleben.de  
Insert length: 612 Std Error: 0.00  
Plate: 1 row: E column: 24  
Seq primer: T7

## FEATURES

Location/Qualifiers

1. 612  
/organism="Hordelum vulgare subsp. vulgare"  
/mol\_type="mRNA"  
/cultivar="Dairke"  
/db\_xref="GABI:557881"  
/clone="HL01B24"  
/issue\_type="embryo, scutellum und aleuron"  
/dev\_stage="seeds, 12 h and 62 h after imbibition"  
/lab\_host="DH10B"  
/clone\_lib="HL"  
/note="Vector: pSPORT1; Site\_1: SalI (5'-end of cDNA); Site\_2: NotI (3'-end of cDNA); average insert size is 1.6-1.7 kb, library was created by RZPD, Heidelberg (Dr. Bernhard Korn, Im Neuenheimer Feld 506, D-69120 Heidelberg Germany, <http://www.rzpd.de>)"

BASE COUNT 114 a 187 c 187 g 124 t

## ORIGIN

## Alignment Scores:

Pred. No.: 5.9e-05 Length: 612  
Score: 146.00 Matches: 47  
Percent Similarity: 42.67% Conservative: 17  
Best Local Similarity: 31.33% Mismatches: 58  
Query Match: 5.56% Indels: 28  
Gaps: 7

US-10-088-045-2 (1-502) x CB882241 (1-612)

QY 93 LysValMetArgLeuAlaSpolyValGluTyrGlnPheTyrThr----- 107  
DB 168 AAGGCCAGGCACCTCAAGTGGAGGACCATGTTCTGCGCGGACCTGCGAGAG 227  
QY 108 -----PheGlyGlyGlnValProGlyGlnMetLeuArgValArgGluGly 122  
DB 228 AAGGCTTCATGGCATCAAGCGCCAGTTCCTCCCGGCCACCATCCGGCCAGCGCG 287  
QY 123 AspThrIleGluValGlnPheSerAsn-----HisProAspSerIleMetProHisAsn 140  
DB 288 GACACATGCTGTGAGCTCAAGAACGGGCTGCACACCGAGGGGTGTCATCCAC--- 344  
QY 141 ValAspPheHisAlaAlaThrGlyProGly-----GlyGlyAlaGluAla 155  
DB 345 -----TGGCAGCGGTCAGACAGATTGGACACCGTGGCGGATGGCAGCGCGCATC 398  
QY 156 SerPheThrAla-----ProGlyHisThrSerThrPheSerPheValAlaLeuGlnPro 173  
DB 399 TCCCAATGGCCATCAACCCCGAGAAACCTTCACTTACCGATTGTCTCGACACAGCGC 458  
QY 174 GlyLeuTyrValTyrHisCysAlaValAlaProValGlyMetHisIleAlaSerGlyMet 193  
DB 459 GGGACATATTTTACCAT-----GGGCACTACGCGATGACAGAGCGCGAGCGCTG 509  
QY 194 TyrGlyLeuIleLeuValGluProLysGluGlyLeuPro-----LysValAspLys 210  
DB 510 TACGCTCTCTGATGTGATGTGGACATGGGAGAGAACCCGTTCAAGTATGACGGC 569  
QY 211 GluTyrTyrValMetGlnGlyAspPheTyr 220  
DB 570 GAGCTGAACCTGTCTCTCAGCAGCTGTGAC 599

## RESULT 16

CAS53053 501 bp mRNA linear EST 19-NOV-2002  
LOCUS C0866F05-5N NIA Mouse Blastocyst cDNA Library (Long) Mus musculus  
DEFINITION CDNA clone NIA:C0866F05 IMAGE:30031360 5', mRNA sequence.  
ACCESSION CAS53053

VERSION CAS53053.1 GI:25097296  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 501)  
Piao, Y., Kargul, G.J., Dudekula, D.B., Qian, Y., Tanaka, T., Luo, A. and

Authors: Ko, M.S.H.  
TITLE Systematic Analyses of NIA Mouse Blastocyst cDNA Library (Long)  
JOURNAL Unpublished  
COMMENT Contact: Dawood B. Dudekula  
Laboratory of Genetics  
National Institute on Aging/National Institutes of Health  
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA  
Email: [cdna@gen.gsc.nia.nih.gov](mailto:cdna@gen.gsc.nia.nih.gov)  
Plate: C0866 Row: F Column: 05  
Seq primer: M13 Reverse  
High quality sequence stop: 501  
POLYA=No.

## FEATURES

source

Location/Qualifiers

1. 501  
/organism="Mus musculus"

/mol\_type="mRNA"

/strain="C57BL/6J"

/db\_xref="nlaEST:C0866F05-5N"

/db\_xref="taxon:10090"

/clone="NIA:C0866F05 IMAGE:30031360"

/issue\_type="Blastocyst"

/dev\_stage="3.5-dpc"

/lab\_host="DH10B"

/clone\_lib="NIA Mouse Blastocyst cDNA Library (Long)"

/note="Vector: pSPORT1 (Invitrogen); Site\_1: SalI; Site\_2: NotI. Mouse cDNA project by the Laboratory of Genetics, National Institute on Aging (NIA), Intramural Research Program, NIH (<http://19sun.grc.nia.nih.gov/cDNA>). This is a long-transcript enriched cDNA library (Ref. Genome Res. 11: 1553-1558 (2001). [PMID: 11544191]). Total RNAs were extracted from a pool of 20 Blastocysts. Double-stranded cDNAs were synthesized with an Oligo(dT) primer

(Invitrogen): 5'-DCACTAGTTAGATCGAGCGCGCCCTTTTCTTTTCTTTT-3' from 0.2 ug of total RNA, treated with T4 DNA polymerase, and purified by ethanol-precipitation. The cDNAs were ligated to lone-linker IL-Sal4, purified by phenol/chloroform, and separated from free linkers by Centricon 100. Then, the cDNAs were amplified by long-range high fidelity PCR using Ex Taq polymerase (Takara) with a primer Sal4-S. The products were purified by phenol/chloroform and Centricon 100. The cDNAs were digested with SalI and NotI enzymes and cloned into SalI/NotI site of pSPORT1 plasmid vector. The DH10B E. coli host was transformed with the ligation mixture by the standard chemical method. The average insert size is about 2.2 kb. The library was constructed by Yulan Piao (NIA)."

BASE COUNT 98 a 152 c 146 g 105 t

## ORIGIN

Alignment Scores:  
Pred. No.: 5.06e-05 Length: 501  
Score: 145.50 Matches: 44  
Percent Similarity: 43.11% Conservative: 28  
Best Local Similarity: 26.35% Mismatches: 66  
Query Match: 5.54% Indels: 29  
DB: 14 Gaps: 8

US-10-088-045-2 (1-502) x CAS53053 (1-501)

QY 60 ProValIleAspAlaIleValThrHisAlaProGluValProProProValAspArgAsp 79  
DB 498 CCGATGACGACAAAG--GGCAGCATGACGCCACCGTTGATCCACGCTCCGCTCCGAC 442  
QY 80 HisProAlaLysValValVal-----LysMetGlu----- 89





ACCESSION CASS1823  
 VERSION CASS1823.1 GI:25095973  
 KEYWORDS EST.  
 SOURCE Mus musculus  
 ORGANISM Mus musculus (house mouse)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 515)  
 Piao, Y., Kargul, G.J., Dudekula, D.B., Qian, Y., Tanaka, T., Luo, A. and Ko, M.S.H.  
 TITLE Systematic Analyses of NIA Mouse Blastocyst cDNA Library (Long)  
 JOURNAL Unpublished  
 COMMENT Contact: Dawood B. Dudekula  
 Laboratory of Genetics  
 National Institute on Aging/National Institutes of Health  
 333 Caswell Drive, Suite 4000, Baltimore, MD 21224-6820, USA  
 Email: cdna@igsun.grc.nia.nih.gov  
 Plate: C0861 row: A column: 06  
 Seq primer: M13 Reverse  
 High quality sequence stop: 515  
 POLYA=No.

# FEATURES

source location/Qualifiers  
 1..515  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="taeST:C0849A06-5N"  
 /db\_xref="taxon:10090"  
 /clone="NIA:C0849A06 IMAGE:30029669"  
 /issue\_type="Blastocyst"  
 /dev\_stage="3.5-dpc"  
 /lab\_host="DH10B"  
 /clone\_lib="NIA Mouse Blastocyst cDNA Library (Long)"  
 /note="Vector: pSPORT (Invitrogen); Site 1: SalI; Site 2: NotI. Mouse cDNA project by the Laboratory of Genetics, National Institute on Aging (NIA), Intramural Research Program, NIH (http://igsun.grc.nia.nih.gov/cDNA). This is a long-transcript enriched cDNA library (Ref. Genome Res. 11: 1553-1558 (2001). [PMID: 11544199]). Total RNAs were extracted from a pool of 20 Blastocysts. Double-stranded cDNAs were synthesized with an Oligo(dt) primer [Invitrogen].  
 5'-pGACTGATCTGATCGCGAGCGCGCCCTTTTCTTTT-3' from 0.2 ug of total RNA, treated with T4 DNA polymerase, and purified by ethanol-precipitation. The cDNAs were ligated to lone-linker L1-SalI, purified by phenol/chloroform, and separated from free linkers by Centricon 100. Then, the cDNAs were amplified by long-range high fidelity PCR using Ex Taq polymerase (Takara) with a primer SalI-S. The products were purified by phenol/chloroform and Centricon 100. The cDNAs were digested with SalI and NotI enzymes and cloned into SalI/NotI site of pSPORT plasmid vector. The DH10B E. coli host was transformed with the ligation mixture by the standard chemical method. The average insert size is about 2.2 kb. The library was constructed by Yulan Piao (NIA)."  
 BASE COUNT 100 a 155 c 151 g 109 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 5,26e-05 Length: 515  
 Score: 145.50 Matches: 45  
 Percent Similarity: 43.20% Conservative: 28  
 Best Local Similarity: 26.63% Mismatches: 67  
 Query Match: 5.54% Indels: 29  
 DB: 14 Gaps: 8

US-10-088-045-2 (1-502) x CASS1823 (1-515)

QY 58 GUAUUPROVALIILASAPLALIEVALTHRHISALAPROGLIUALVAPROPROVALIASP 77  
 DB 504 GAAGCGCATCATGACAAAG--CGACGATGACGCCCGGTGCATCCACGTCGGAT 448

QY 78 ARGAPRHISPROVALIYVALIVAL-----LysMetGlu--- 89  
 DB 447 CCCGACTATCGGCGCGGTGACGCTGAACGGCTGCTGTCATGGCGCATGAACGGC 388  
 QY 90 -----ThrValGluValMetArg---LeuAlaAspGlyVal 101  
 DB 387 GACTGGAAGAGATTTCACCTGTGCGCAACCTGTGGCGCAATTGGCCAGAGCATG 328  
 QY 102 GUTYRGINPHEPTTRPHEGLYGLYGINVALPROGLYGLIMETILARGLVALRGL 121  
 DB 327 AAGGCTCATCTTGTGGGTTCACACGCCATCCCGGTCCACCATTTGAGCAGTAGNA 268  
 QY 122 GLYAPTRHILIEGLUALGLNPHESERASNHSIPROASPSERLYMETPROHISAN--- 140  
 DB 267 GCGCAGCAAGCTCCGATCTTCGTCCACCAAC-----AAATGGCGGAGCATACC 220  
 QY 141 ---VALAPRHEHIALALATHRGLYPROGLY-----GlyGlyAlaGlu 154  
 DB 219 ACCGTACATTTGGCAGCGGTCTTTCGCGAGCGGTATGACGGGTCGCGCTCAAC 160  
 QY 155 ALASERPHETHRALAPROGLYHISHTSERTHRPHESERPHELVALLEUGLNPROGLY 174  
 DB 159 CAGCTCACATCAAACTGGCGAGACCTTCCTTCGAGTTCCAGATGAGCATACGGCG 100  
 QY 175 LEUTYRVALTYRHISCYALAVALAPROVALGLYMETHSILALASNGLYMETTYR 194  
 DB 99 ACGTTATGTACCAAC---CGCAGCTCGACGCAATGATGCGATGCGATGCGCATGATG 43  
 QY 195 GYLEUITLEUVALGLUPROLYSGLU 203  
 DB 42 GCGCATGATGTGTGCATCCACGTCAT 16

RESULT 19  
 LOCUS CASS2719/515 bp mRNA linear EST 19-NOV-2002  
 DEFINITION C0861B10-5N NIA Mouse Blastocyst cDNA Library (Long) Mus musculus  
 ACCESSION CASS2719  
 VERSION CASS2719.1 GI:25096939  
 KEYWORDS EST.

REFERENCE  
 AUTHORS Piao, Y., Kargul, G.J., Dudekula, D.B., Qian, Y., Tanaka, T., Luo, A. and Ko, M.S.H.  
 TITLE Systematic Analyses of NIA Mouse Blastocyst cDNA Library (Long)  
 JOURNAL Unpublished  
 COMMENT Contact: Dawood B. Dudekula  
 Laboratory of Genetics  
 National Institute on Aging/National Institutes of Health  
 333 Caswell Drive, Suite 4000, Baltimore, MD 21224-6820, USA  
 Email: cdna@igsun.grc.nia.nih.gov  
 Plate: C0861 row: B column: 10  
 Seq primer: M13 Reverse  
 High quality sequence stop: 515  
 POLYA=No.

# FEATURES

source location/Qualifiers  
 1..515  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="taeST:C0861B10-5N"  
 /db\_xref="taxon:10090"  
 /clone="NIA:C0861B10 IMAGE:30030837"  
 /issue\_type="Blastocyst"  
 /dev\_stage="3.5-dpc"  
 /lab\_host="DH10B"  
 /clone\_lib="NIA Mouse Blastocyst cDNA Library (Long)"  
 /note="Vector: pSPORT (Invitrogen); Site 1: SalI; Site 2: NotI. Mouse cDNA project by the Laboratory of Genetics, National Institute on Aging (NIA), Intramural Research





Pred. No.:	8.18e-05	length:	59e
Score:	144.50	Matches:	50
Percent Similarity:	40.10%	Conservative:	29
Best Local Similarity:	25.38%	Mismatches:	81
Query Match:	5.50%	Indels:	38
DB:	29	Gaps:	5

US-10-088-045-2 (1-502) X CC016516 (1-596)

OY 31 AlaGlnProLySerSerThrValAlaPheIleAlaLeuTySthrAlaAsnAlaAsnAla 50  
Db 25 TCATCCCCCGGTAAAGGATGCACAGCGAAGTCTCTCCACCAATCCTGGCGGTGAGAAGCG 84  
OY 51 AlaserGInGuHISGLNGlyGlndeUroValIleAepAlaIleValThnHislaPro 70  
Db 85 GAGCAGCGCCGGCAGCGTGAGACACTGACCTCCGAAGCGCATTTGACCCTGGCGCT 144  
OY 71 GluValProProProValAspArgAspHisProAlaLysValValValLysMeCluThr 90  
Db 145 -----ACCAACG 150  
OY 91 ValGluLysValMetArgLeuAlaMetGlyValGluTyrgInPheTrpThreGly 110  
Db 151 GTAGAG-----CTCGGTGCCGGCTCTCCGCCACCACTGGCGCTTGCTCT 198  
OY 111 GluValProGlyGlnMetCileArgValArgrGluGlyAspThrIleGluValGlnPheser 130  
Db 199 GCGGTACCCGGGCCGGGAATCGGTCTTGTAAGGCCAGCGCTACGGCTTCAGGTGAG 258  
OY 131 AsnHisProAspSerLysMetProHisAsnValAspPheHisAlaIle----- 146  
Db 259 AAC-----CGCTTCCGACCGGCACTGCTCTCACTG- GCATGTGCTTAGCAAT 305  
OY 147 -----ThnGlyProGlyGlyValAlaGluAlaSerPheThrAlaProGlyHisThr 163  
Db 306 ACCGACAGCATGATGGGGGTGCAGGGTTGACGGCATGTCAGCATTTCCCCCGGACATCTC 365  
OY 164 SerThr-PheSerPheLysAlaLeuGlnProGlyLeuTyrgValTyrgHisCySAlaValAla 183  
Db 366 ACAACGGGTACGAGTTCATCGTCCCAGCAGCGGGCACGATCGTGTATCAC-----CC 416  
OY 183 AProValGelwethisIleAlaAsnGlyMetTyrgLysLeuIleLeuValGluProLysGel 203  
Db 417 CCACGCTCGGGGTACAGCTTCGACCGCGGGGTGTACCGCCACTGATCATCATGACGCCGA 476  
OY 203 uGlyLeuProLysValAspLysGluTyrgTyrgLysMetGlnGlyAspPhe 219  
Db 477 CGAGCGCACCGCTTACGACGAGAAATGGATCTCGTCTTCGATGCTGG 525

[illegible]

High quality sequence stop: 360	
POLYA=No.	
Location/Qualifiers	
1..360	
source	
FEATURES	

High quality sequence stop: 360  
POLYA=No.

```

/organism="mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taeSTR:C0836D10-5N"
/db_xref="taxon:10090"
/cclone="NTA:C0836D10 IMAGE:30028461"
/clisue_type="Blastocyst"
/dev_stage="3.5-dpc"
/lab_host="DH10B"
/clone_1ib="NTA Mouse Blastocyst cDNA Library (Long)"
/clone_vector:psPORT1 (Invitrogen); Site 1: SalI; Site 2:
NotI; Mouse cDNA project by the Laboratory of Genetics,
National Institute on Aging (NIA), Intramural Research
Program, NIH (http://lgun.gic.nia.nih.gov/cDNA). This is
a long-transcript enriched cDNA library (Ref. Genome Res.
11: 1553-1558 (2001). [PMID: 11544191]). Total RNAs were
extracted from a pool of 20 Blastocysts. Double-stranded
cDNAs were synthesized with an oligo(dT) primer
[Invitrogen:
5'-ggaCTAGTCTTCAATCGGAGCGGCCGCCCTTTTTTT-3'] from
0.2 ug of total RNA, treated with T4 DNA polymerase, and
purified by ethanol-precipitation. The cDNAs were ligated
to lone-linker lp-Sal4, purified by phenol/chloroform, and
separated from free linkers by Centricon 100. Then, the
cDNAs were amplified by long-range high fidelity PCR using
Ex Taq polymerase (Takara) with a primer Sal4-S. The
products were purified by phenol/chloroform and Centricon
100. The cDNAs were digested with SalI and NotI enzymes
and cloned into SalI/NotI site of psPORT1 plasmid vector.
The DH10B E. coli host was transformed with the ligation
mixture by the standard chemical method. The average
insert size is about 2.2 kb. The library was constructed
by Yulan Piao (NIA)."
BASE COUNT      74 a      105 c      102 g      79 t
ORIGIN
Alignment Scores:
Pred. No.:      5,21e-05      Length:      360
Score:          143.50      Matches:      35
Percent Similarity: 47.50%      Conservative: 22
Best local Similarity: 29.17%      Mismatches:   50
Query Match:     5.46%      Indels:      13
DB:             14      Gaps:         4

```

```

Qy      92 GILuysValmetArg---LeuAlaaspGlyValaIGuIyGlnPhehrPthPheGly 110
Db      360 GAACCTGGTGGTCCGGGAATTTTCCGAAGGCATGAAGCTCATCTCTGGGGTTTCAACGGC 302
Qy      111 GlnValProGlyIGlMetCilleaArgValaIagGInGlyAspThrIleGluValGlnPheSer 130
Db      300 CAATCCCGGGGTCCACCATTTGAGGCACATGAAGAGGCAAGCTCCGTATCTTCGTACCC 241
Qy      131 AsnHisProAspSerSellysMetProHisAsnValaIaspPheHis-----144
Db      240 AAC-----AAATTCGCGGAGCATACCAACCGTACATGTGGCATGGCGTCTTTTG 193
Qy      145 ---AlaAlaThrGlyProGlyGlyGlyValaGluAlaSerPheThrAlaProGlyHisThr 165
Db      192 CCGACCGCGTATGGGGGGGGTTCGGCGGCTTCAACACAGCTTCACATCAAACTGGCGAGACC 133
Qy      164 SerThrPheSerPheLysAlaIaLeuGlnProGlyLeuTrValTrpHisCysAlaValAla 183
Db      132 TTCGTCCTACAGATCTCCAGATGAAGCATACGGGAGCCTTATGTGACAC---CCGCACTCC 76
Qy      184 ProValIGlyMetCisIleAlaAsnGlyMetTrGlyLeuIleLeuValaGluProLysGlu 203
Db      75 GACGAATGTGTGCAGATGGCGCATGGCCATGATGGGATATGTTGGTGCATCCACCGTAT 16

```



## ORIGIN

## Alignment Scores:

Pred. No.: 0.000141 Length: 623  
 Score: 142.50 Matches: 47  
 Percent Similarity: 44.63% Conservative: 32  
 Best Local Similarity: 26.55% Mismatches: 59  
 Query Match: 5.43% Indels: 39  
 DB: 9 Gaps: 10

US-10-088-045-2 (1-502) x AV822649 (1-623)

```

OY 101 ValGluTyrGlnPheThr-----ThrpheGlyGly 110
    |||||.....|
DB 108 GTGGAGTACAGATTGTCGCCGACACGAAAGAGCGCGCTTATGACCGTCAACGGC 167
    |||||.....|
OY 111 GlnValProGlyGlnMetIleArgValArgGluGlyAspThrIleGluValGlnPheSer 130
    |||||.....|
DB 168 GAGTTTCTGTCGCCACCATTAAGCTTCGCCGAGACACCATCGTCGTCATCTCACC 227
    |||||.....|
OY 131 AennIaProAspSerLysMetProHisAsn-----ValAspPheHisAla----- 145
    |||||.....|
DB 228 AAC-----AACTCACACCGAAGCGCTTGTATCATTCATTCGATGGATCCGT 275
    |||||.....|
OY 146 -----AlaThrGlyProGlyGlyValGluValAspPheThrAla 159
    |||||.....|
DB 276 CAGTTCCGAGATCCATGGCGAGATGAGACAGACAGAGTTACTCAATGCCCATTAAC--- 332
    |||||.....|
OY 160 ProGlyHisThrSerThrPheSerPheLysAlaLeuGlnProGlyLeuTyrValTyrHis 179
    |||||.....|
DB 333 CCGGAGAGACTTTTACCTACATTCATCTGTAAGACCGGAAACACATTTCTACAT 392
    |||||.....|
OY 180 CysAlaValAlaProValGlyMetHisIleAlaAsnGlyMetTyrGlyLeuIleVal 199
    |||||.....|
DB 393 -----GACACATATGGCATGAGATCAGATCAGCTGCTATACGATCGTGAATGTG 443
    |||||.....|
OY 200 GluProLysGlnGlyLeuPro-----LysValAspLysGlyTyrValMetGln 216
    |||||.....|
DB 444 GACGTGGCTTAAGAAAGACGACAGATTTGAGATACATGTTAATCTTAACTC 503
    |||||.....|
OY 217 GlyAspPheTyrThrLysGly---LysTyrGlyGlnGlnLeuGlnProPheAspMet 235
    |||||.....|
DB 504 AGTGAATGTCGATGATGAGCTATTCCTCCCAAGAACTGCTCTTCT----- 551
    |||||.....|
OY 236 GluLysAlaIleArg-----GluAspAlaGluTyrValValPheAsnGly 250
    |||||.....|
DB 552 TCCAAACCTATGCGCTGATCGGTGAAGCTCAGAGCATATTGATTAATGGG 602
    |||||.....|

RESULT 26
CA483187 691 bp mRNA linear EST 14-NOV-2002
LOCUS LUP12006H1LR LUP12 Linum usitatissimum cDNA clone LUP12006H1LR,
DEFINITION mRNA sequence.
ACCESSION CA483187
VERSION CA483187.1 GI:24961114
KEYWORDS EST.
SOURCE Linum usitatissimum (Flax)
ORGANISM Linum usitatissimum
    BkaraYota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
    Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
    1; eustosida I; Malpighiales; Linaceae; Linum.
    1 (bases 1 to 691)

REFERENCE
AUTHORS Cloutier,S.
TITLE Flax genomics: ESTs from developing bolls
JOURNAL Unpublished
COMMENT Contact: Dr. Sylvie Cloutier
    Cereal Research Centre, Agriculture and Agri-Food Canada
    195 Dafoe Rd, Winnipeg, MB, Canada R3T 2M9
    Tel: (204) 983-2340
    Fax: (204) 983-4604
    Email: scloutier@em.agr.ca
    was cloned using a SalI adapter. The insert is located between the
    NotI and MluI sites. Not all sequences generated with reverse
  
```

## FEATURES

## source

primer were from the 5' end (same with forward primer and 3' end).  
 Average insert size is 2.07 kb  
 Plate: 006 row: H column: 11  
 Seq primer: M13 Reverse  
 Location/Qualifiers

```

1..691
/organism="Linum usitatissimum"
/mol_type="mRNA"
/cultivar="Mc Mcduff"
/db_xref="taxon:4006"
/clone="LUP12006H1R"
/cisue_type="Bolls"
/dev_stage="12 days after flowering"
/lab_host="EMDH108-70NA"
/clone_id="LUP12"
/Note="Vector: pCMVSPORTe.1 (Invitrogen Corp.); Site_1:
NotI; Site_2: MluI; mRNA obtained from bolls 12 days after
flowering"

```

BASE COUNT 170 a 174 c 195 g 152 t  
 ORIGIN

## Alignment Scores:

Pred. No.: 0.000162 Length: 691  
 Score: 142.50 Matches: 49  
 Percent Similarity: 41.85% Conservative: 28  
 Best Local Similarity: 26.63% Mismatches: 76  
 Query Match: 5.43% Indels: 31  
 DB: 14 Gaps: 8

US-10-088-045-2 (1-502) x CA483187 (1-691)

```

OY 78 ArgAspHisProAlaLysValValValLysMetGluThrValGlu-----Lys 93
    |||||.....|
DB 108 CCGGTTCAATGTTGATGATGATGATGTCGCGCGCTTCAAGTTCTGCGGGAAA 167
    |||||.....|
OY 94 ValMetArgLeuAlaAspLysValGluTyrGlnPheThr----- 107
    |||||.....|
DB 168 ACTAGGCAATTCMAATGGACGTGGAGTACATGTTCTGTGACCGGACGAGAGACAT 227
    |||||.....|
OY 108 -----PheGlyGlnValProGlyGlnMetIleArgValArgGluGlyAsp 123
    |||||.....|
DB 228 ACTGTCATGGATTAACACGCGCATGTTCCGGGTCCGACATTAAGGCCAAAGTCCGAC 287
    |||||.....|
OY 124 ThrIleGluValGlnPheSerAsn-----HisProAspSerLysMetProHisAsnVal 141
    |||||.....|
DB 288 ATATATCAAGTGAAGCTCAGCAAGATCTCCATACGAGAGGATGTTATTCAGTGGCAC 347
    |||||.....|
OY 142 AspPheHisAlaAlaThrGlyProGlyGlyValGluValAspPheThr----- 158
    |||||.....|
DB 348 GGCATCAGACAGAAAGGACATCCATGCGCAGATGGACAGATTCATCTCACAGTCCGCC 407
    |||||.....|
OY 159 ---AlaProGlyHisThrSerThrPheSerPheLysAlaLeuGlnProGlyLeuTyrVal 177
    |||||.....|
DB 408 ATGCTCTCGTGTGAACCTTCCCTTACAGTTCACTGTGACACAGCAGGACATATTTC 467
    |||||.....|
OY 178 TyrHisCysAlaValAlaProValGlyMetHisIleAlaAsnGlyMetTyrGlyLeuIle 197
    |||||.....|
DB 468 TACCAC-----GCCATTACGGAGATGCAAGGTCCAGCAGGCGCTATACGTTTCATTG 518
    |||||.....|
OY 198 LeuValGluProLysGlnGlyLeuProLys-----ValAspLysGluTyr 212
    |||||.....|
DB 519 ATAGTGAAGCCGGAATATATGCGGGAGAAAGATTCATTCATTAAGATGCGAGATTC 578
    |||||.....|
OY 213 TyrValMetGlnLysPheTyrThrLysGlyLysTyrGlyGluGln---GlyLeuGln 231
    |||||.....|
DB 579 AACCTCTGCTCAGCAGCTGTCGTCATTAAGTTCACAGACCAAAAGTGGCTCTCT 638
    |||||.....|
OY 232 -----ProPhe 233
    |||||.....|
DB 639 TCCCTCCCTTT 650
    |||||.....|

RESULT 27
AY079863/c
  
```

LOCUS AY079863 752 bp DNA linear GSS 06-NOV-2002  
 DEFINITION AY079863 Serpins Pier (La Jolla, CA) uncultured virus community  
 uncultured marine virus genomic clone SIO51p10FIL, genomic survey  
 sequence.  
 ACCESSION AY079863  
 VERSION AY079863.1 GI:24744722  
 KEYWORDS GSS.  
 SOURCE uncultured marine virus  
 ORGANISM uncultured marine virus  
 REFERENCES Viruses; environmental samples.  
 1 (bases 1 to 752)  
 AUTHORS Breitbart, M., Salomon, P., Andreesen, B., Mahafey, J.M., Segall, A.M.,  
 Mead, D., Azam, F. and Rohwer, F.  
 TITLE Genomic analysis of uncultured marine viral communities  
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (22), 14250-14255 (2002)  
 MEDLINE 22294988  
 PUBMED 12384570  
 COMMENT Contact: Rohwer F  
 Biology Dept.  
 San Diego State University  
 5500 Campanile Dr, San Diego, CA 92102, USA  
 Tel: 6195941336  
 Fax: 619595676  
 Email: forest@unstroke.sdsu.edu  
 Class: Shotgun.  
 FEATURES  
 source  
 1. 752  
 /organism="uncultured marine virus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:186617"  
 /clone="SIO51p10FIL"  
 /clone\_lib="Serpins Pier (La Jolla, CA) uncultured virus  
 community"  
 /note="Marine viruses were isolated from 200 liters of  
 surface seawater using a combination of differential  
 filtration and density-dependent gradient centrifugation.  
 Linker-amplified shotgun libraries were constructed by  
 randomly shearing the total marine viral community DNA,  
 end-repairing, ligating deDNA linkers to the ends, and  
 amplifying the fragments using Vent DNA polymerase. The  
 resulting fragments were ligated into the pSMART vector  
 and electroporated into Wc12 cells (Lucigen; Middleton, WI  
 )"  
 BASE COUNT 117 a 225 c 261 g 148 t 1 others  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 0.000206 Length: 752  
 Score: 142.00 Matches: 49  
 Percent Similarity: 52.07% Conservative: 14  
 Best Local Similarity: 40.50% Mismatches: 41  
 Query Match: 5.41% Indels: 19  
 DB: 28 Gaps: 6  
 US-10-088-045-2 (1-502) x AY079863 (1-752)  
 QY 388 ProAlaAsnLeuGlnGluGln-----IleLysAlaGlyLysAlaThrTyrAspSer 404  
 DB 752 CCNAAACAGATTGGGAGCGAGCGCGCCCTGATGCCCCGGGCGAGAGGCTTATGCGAAC 693  
 QY 405 AAcCYsAlaAlaCyHisGlnProAspGlyLysGlyValProAsnAlaPheProProla 424  
 DB 692 AACCTGGCGGCTGTGCAC-AAGCCCGAGCGGCAAGGGCG--CCGGCCGCTTCCCCGGCCCTC 636  
 QY 425 AlaAsnSerAspTyrLeuAsnAla---AspHisAlaArgAlaAlaSerIleValAlaAsn 443  
 DB 635 GATGCTCCGCCGCTGTCGACGACGACGACGACGACGACGACGACGACGACGACGAC 576  
 QY 444 GlyLeuSerGlyLysIleThrValaAsnGlyLeuSerValaMetProAlaIle 463  
 DB 575 GCGCGC-----AACACCAACACG-----ATGCTGCTGCG 546  
 QY 464 -----AlaLeuSerAspGlnGlnIleAlaAsnValIleThrTyrThrLeuAsnSerPhe 481

DB 545 AAGGCGAACCTCTCGGACGACGATCGCGCGCCGATACGACGAGAACAGCTCG 486  
 QY 482 GlyAsnLysGlyGln---LeuSerAlaAspAspValAlaLysAlaLysThrLys 500  
 DB 485 TCCACACAGACCGGACGACGACGACGACGACGACGACGACGACGACGACGAC 426  
 QY 501 Pro 501  
 DB 425 CCC 423  
 RESULT 28  
 CAS54547/c  
 LOCUS  
 DEFINITION C0887A01-5N NIA Mouse Blastocyst cDNA Library (Long) Mus musculus  
 cDNA clone NIA:C0887A01 IMAGE:30033312 5', mRNA sequence.  
 ACCESSION CAS54547.1 GI:25098838  
 VERSION  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 502)  
 Piao, Y., Kargul, G.J., Dudekula, D.B., Qian, Y., Tanaka, T., Luo, A. and  
 Ko, M.S.H.  
 TITLE Systematic Analyses of NIA Mouse Blastocyst cDNA Library (Long)  
 JOURNAL Unpublished  
 COMMENT Contact: Dawood B. Dudekula  
 Laboratory of Genetics  
 National Institute on Aging/National Institutes of Health  
 333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA  
 Email: cdaa@igsun.grc.nia.nih.gov  
 Plate: C0887 row: A column: 01  
 Seq primer: M13 Reverse  
 High quality sequence stop: 502  
 POLYA=No.  
 FEATURES  
 source  
 1. 502  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="taeST:C0887A01-5N"  
 /db\_xref="taxon:10090"  
 /clone="NIA:C0887A01 IMAGE:30033312"  
 /issue\_type="Blastocyst"  
 /dev\_stage="3.5-dpc"  
 /lab\_host="DH10B"  
 /clone\_lib="NIA Mouse Blastocyst cDNA Library (Long)"  
 /note="Vector: pSPORT1 (Invitrogen); Site 1: SalI; Site 2:  
 NotI. Mouse cDNA project by the Laboratory of Genetics,  
 National Institute on Aging (NIA), Intramural Research  
 Program, NIH (http://igsun.grc.nia.nih.gov/cDNA). This is  
 a long-transcript enriched cDNA library (Ref. Genome Res.  
 11: 1553-1558 (2001)). [PMID: 11544191]. Total RNAs were  
 extracted from a pool of 20 Blastocysts. Double-stranded  
 cDNAs were synthesized with an Oligo(dT) primer  
 [Invitrogen:  
 5'-pGACTAGTTCAGATCGGAGCGCGCCCTTTT-3'] from  
 0.2 ug of total RNA, treated with T4 DNA polymerase, and  
 purified by ethanol-precipitation. The cDNAs were ligated  
 to one-linker ltr-SalI, purified by phenol/chloroform, and  
 separated from free linkers by Centricon 100. Then, the  
 cDNAs were amplified by long-range high fidelity PCR using  
 Ex Taq polymerase (Takara) with a primer SalI-S. The  
 products were purified by phenol/chloroform and Centricon  
 100. The cDNAs were digested with SalI and NotI enzymes  
 and cloned into SalI/NotI site of pSPORT1 plasmid vector.  
 The DH10B E. coli host was transformed with the ligation  
 mixture by the standard chemical method. The average  
 insert size is about 2.2 kb. The library was constructed  
 by Yulan Piao (NIA)."  
 BASE COUNT 99 a 152 c 145 g 106 t

## ORIGIN

## Alignment Scores:

Pred. No.:	0.000133	Length:	502
Score:	141.50	Matches:	44
Percent Similarity:	43.11%	Conservative:	28
Best Local Similarity:	26.35%	Mismatches:	66
Query Match:	5.39%	Indels:	29
DB:	14	Gaps:	8

US-10-088-045-2 (1-502) x CA554547 (1-502)

```

QY 60 ProvaliLeAspAlaLeValThrHisAlaProGluValProProProValAspArgAsp 79
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 498 CCGATATGACACAG--GCGACGATGACAGCCACCGTGCATCCACGCCGCCGAC 442
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 80 HisProAlaLeValValValValValValValValValValValValValValVal 89
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 441 TATCGCCGGTGTGACGCGTGAACGGCTGCTGTCATGCGGCATGAACGGCGACTGG 382
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 90 -----ThrValGluValValMetArg---LeuAlaAspGluValGluTyr 103
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 381 AAGAGATTACCTGCTGCGCCGACCACTGTGTCGCGCAATTTCGCAAGCATGAAGCT 322
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 104 GlnPheThrPheGluGluGluValProGluGluMetIleArgValArgGluGluAsp 123
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 321 CATCTGGGGTTACACGGCCAAATCCCGGCTCCACCATTAAGCAGTGAAGAGCGAC 262
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 124 ThrIleGluValGlnPheSerAsnHisProAspSerIleMetProHisAsn-----Val 141
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 261 AAGCTCGATCTTCGTCACCAAC-----AAATTGCGGAGCATACCACTGTA 214
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 142 AspPheHisAlaAlaThrGlyProGly-----GlyValAlaGluAlaSer 156
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 213 CATTCGATGCGCGCTGTTTCCGACGCGGTATGACGCGGCTCGCGCTCAACCAAGCT 154
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 157 PheThrAlaProGluHisThrSerThrPheSerPheValAlaLeuGlnProGluLeuTyr 176
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 153 CACATCAAACTGGCGACAGCTTCGTCATGAGTTTCAGATGAAGCATAGCGGAGCTTT 94
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 177 ValTyrHisCysAlaValAlaProValGluMetHisIleAlaAsnGluMetTyrGluLeu 196
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 93 ATGTACACAC---CCGCACTCCGACGAATATGTCAGATGCGCATGATGAGGACATG 37
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 197 IleLeuValGluProGlyGlu 203
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 36 ATGTGTGTCATCCAGCTGAT 16
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

## RESULT 29

AV826062 600 bp mRNA linear EST 01-APR-2002  
 LOCUS AV826062 RAEFL8 Arabidopsis thaliana cDNA clone RAEFL08-08-024 5',  
 DEFINITION mRNA sequence.

## ACCESSION

AV826062

## VERSION

AV826062.1 GI:19868122

## KEYWORDS

EST

## SOURCE

Arabidopsis thaliana (thale cress)

## ORGANISM

Arabidopsis thaliana

## REFERENCE

AUTHORS

Seki, M., Narusaka, M., Ishida, J., Kamiya, A., Satou, M., Nakajima, M.,  
 Oono, Y., Sakurai, T., Carninci, P., Kawai, J., Itoh, M., Ishii, Y.,  
 Arakawa, T., Shibata, K., Shingawa, A., Muramatsu, M., Hayashizaki, Y.  
 and Shinozaki, K.  
 Large scale analysis of Arabidopsis full-length cDNA (2002b)

## JOURNAL

COMMENT

Contact: Motoaki Seki  
 Plant Functional Genomics Research Group  
 RIKEN Genomic Sciences Center  
 3-1-1 Koyada, Tsukuba, Ibaraki 305-0074, Japan  
 Tel: 81-298-36-4359

## FEATURES

## source

Location/Qualifiers  
 1..600  
 /organism="Arabidopsis thaliana"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:3702"  
 /clone="RAFL08-08-024"  
 /dev\_stage="rosette plants"  
 /lab\_host="DH10B"  
 /clone\_id="RAFL8"  
 /note="Site 1: BamHI; Site 2: SalI; subjected to  
 dehydration-treated (1, 2, 5, 10, 24 hr)"  
 154 a 137 c 166 g 142 t 1 others

## BASE COUNT

154 a 137 c 166 g 142 t 1 others

## ORIGIN

Alignment Scores:  
 Pred. No.: 0.00017 Length: 600  
 Score: 141.50 Matches: 45  
 Percent Similarity: 43.68% Conservative: 31  
 Best Local Similarity: 25.86% Mismatches: 65  
 Query Match: 5.39% Indels: 33  
 DB: 9 Gaps: 8

US-10-088-045-2 (1-502) x AV826062 (1-600)

```

QY 101 ValGluTyrGlnPheTrp-----ThrPheGluGly 110
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 101 GTGAGTACAAAGTATATGTCGCCGACCTGCAAGAGGCGCGTTATGACCGTCAAGCGC 160
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 111 GlnValProGluGluMetIleArgValArgGluGluAspThrIleGluValGlnPheSer 130
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 161 GAGTTTCTGTGTCACCACTTAAGCTTCGCCGAGACACCATCTGTCATATCTCAC 220
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 131 AsnHisProAspSerIleMetProHisAsn-----ValAspPheHisAla----- 145
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 221 AAC-----AAACTCACCAACCGAAGCGCTTGTGATGATGCGATGGAATCCGT 268
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 146 -----AlaThrGlyProGluGluGluValAlaGluAlaSerPheThrAla 159
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 269 CAGTTCGAAATGCATGCGGACAGATGAGACAGAGAGATTAATCTCAATGCGCAATTAAAC--- 325
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 160 ProGluHisThrSerThrPheSerPheValAlaLeuGlnProGluLeuTyrValTyrHis 179
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 326 CTGAGAGACCTTTTACCTACATTTCTGATGAAAACCGGAAACATTTCTACAT 385
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 180 CysAlaValAlaProValGluMetHisIleAlaAsnGluMetTyrGluLeuVal 199
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 386 -----GACACATGATGCGATGAGATGAGATGAGATGAGATGAGATGAGATGAG 436
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 200 GluProGluGluGluLeuPro-----LysValAspGluGluTyrTyrValMetGln 216
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 437 GACGTGCTAAAGGAAAGAGAGAGATTAAGATGATGATGATGATGATGATGATGATG 496
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 217 GlyAspPheThrThrPheGluGluGluGluGluGluGluGluGluGluGluGluGlu 236
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 497 AGTGAAGTGGTGGATGAGCT---ATTCCCTCCCAAGAACTGGCTTTTTCACAACT 553
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 237 LysAlaIleArgGluAspAlaGluTyrValValPheAsnGly 250
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 554 ATGCGCTGATGCGGGAAGCTCAGAGCATATGTTAATGCGG 595
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

## RESULT 30

BZ560121 1269 bp DNA linear GSS 17-DEC-2002  
 LOCUS BZ560121  
 DEFINITION pac62-164\_2164.x1 pac62-164 Pseudomonas aeruginosa genomic clone





```

QY 53 GINGLHIGLGLGLYLULEUPROVALILEAPRALILEVALTHRHISALAPROGLUVAL 72
DB 171 -----GATGCCAAGGTGCACAC----- 188
QY 73 PROPROVALAARPAARPAARPROVALALYVALVALVALYLMETGLUTHRYALGU 92
DB 189 -----CACACCTGGACATCGGTACACATCAACAAAGTCTCTGCAC 227
QY 93 LYSVALMETARGLEUVALAARGLYVALGLYTRGLINPHEPTRHPHEGLYGLINVAL 112
DB 228 TCGGTCACACAGCTGCG-----GTGACCATCAACCGCGGTC 266
QY 113 PROGLYGLMETLEARGVALARGGLUGLYAARPHILLEGVALGLINPHESEARHIS 132
DB 267 CCGGGGCGACCATCGCGCGACCGGACCGGTGGTGTACCGTCCGCAACAGT 326
QY 133 PROAERSELYMECPROHISAENVALAARPHENIALAALATHRGILYPROGLYGLY--- 151
DB 327 CTGGAGACAGAAACAC---GGCATTCACCTGSCACGGATCCGCGAGGTCCGCGG 383
QY 152 -----GLYALAGUALASERPHETRALAPROGLYHISTHRSERTHR 165
DB 384 TGGGCGGACGCGACCGTCCGCGTCACGAGTGTCCATCTCCCGCGGACACTTCACC 443
QY 166 PHESEPHALYVALALEUGINPROGLYLEUTRYVALYHISGYALVALAALAPROVAL 185
DB 444 TACAGGTGTCTGTCCACCGGACCGCTACTTCTTACAC-----GCCCATCT 494
QY 186 GLYMEHISILEALAEANGLYMETRYGLYLEULELEUVALGLUPROLYSGILYGLY 205
DB 495 GGGATCCAGCGGTGCGCGCGGTCTGACCGGATCTGTGTGTGCGTCCGCGGCGG 554
QY 206 PROLYS-----VALAPLYSGLYTRYTRYVALMETGLINLYAPHERTRYHLYS 222
DB 555 GCGGACCGCTCTCCCTACGACGAGACCGGACCGCTCTCTCATGACGTGTGCGACAG 614
QY 223 GLYLYTRYGLYGLUGLNGLYLEUGINPROPHAEPMELGLYUVALALEARGGLUAP 242
DB 615 AGCGTGTAC-----GAGCAGGCGGTGCGGCTCGG 644
QY 243 ALAGLUTRYVALYALPHEANGLYSERVALGLYALALEU 255
DB 645 TCCGACCGCTCTCTTCTGTGCGGACGACGACGACGCTCTG 683

```

RESULT 32  
 CAS53237 515 bp mRNA linear EST 19-NOV-2002  
 LOCUS C0869G05-5N NIA Mouse Blastocyst cDNA Library (Long) Mus musculus  
 DEFINITION CDNA clone NIA:C0869G05 IMAGE:30031660 5', mRNA sequence.

ACCESSION CAS53237  
 VERSION CAS53237.1 GI:25097478  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 515)  
 AUTHORS Piao, Y., Kargul, G.J., Dudekula, D.B., Qian, Y., Tanaka, T., Luo, A. and Ko, M.S.H.  
 TITLE Systematic Analyses of NIA Mouse Blastocyst cDNA Library (Long)

JOURNAL Unpublished  
 COMMENT Contact: Dawood B. Dudekula  
 Laboratory of Genetics  
 National Institute on Aging/National Institutes of Health  
 333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA  
 Email: cdna@lgsun.irc.nia.nih.gov  
 Plate: C0869 row: G column: 05  
 Seq primer: M13 Reverse  
 High quality sequence stop: 515  
 POLYANo.

FEATURES  
 source Location/Qualifiers  
 1..515

```

/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="nibEST:C0869G05-5N"
/db_xref="taxon:10090"
/clone="NIA:C0869G05 IMAGE:30031660"
/tissue_type="Blastocyst"
/dev_stage="3.5-dpc"
/lab_host="DH10B"
/clone_id="NIA Mouse Blastocyst cDNA Library (Long)"
/notes="Vector: pSPORT1 (Invitrogen); Site 1: SalI; Site 2: NotI; Mouse cDNA project by the Laboratory of Genetics, National Institute on Aging (NIA), Intramural Research Program, NIH (http://lgsun.irc.nia.nih.gov/cDNA). This is a long-transcript enriched cDNA library (Ref. Genome Res. 11: 1553-1558 (2001). [PMID: 11544199]). Total RNAs were extracted from a pool of 20 Blastocysts. Double-stranded cDNAs were synthesized with an Oligo (dT) primer (Invitrogen): 5'-DGACTAGTCTCTGATCGGACGCGCGCCCTTTTCTTTT-3' from 0.2 ug of total RNA, treated with T4 DNA polymerase, and purified by ethanol precipitation. The cDNAs were ligated to lone-linker L1-Sal4, purified by phenol/chloroform, and separated from free linkers by Centricon 100. Then, the cDNAs were amplified by long-range high fidelity PCR using Ex Tag polymerase (Takara) with a primer Sal4-S. The products were purified by phenol/chloroform and Centricon 100. The cDNAs were digested with SalI and NotI enzymes and cloned into SalI/NotI site of pSPORT1 plasmid vector. The DH10B E. coli host was transformed with the ligation mixture by the standard chemical method. The average insert size is about 2.2 kb. The library was constructed by Yulan Piao (NIA)."

```

BASE COUNT 100 a 156 c 151 g 108 t  
 ORIGIN

#### Alignment Scores:

Pred. No.: 0.000176 Length: 515  
 Score: 140.50 Matches: 45  
 Percent Similarity: 42.60% Conservative: 27  
 Best Local Similarity: 26.63% Mismatches: 68  
 Query Match: 5.35% Indels: 29  
 DB: 14 Gaps: 8

US-10-088-045-2 (1-502) x CAS53237 (1-515)

```

QY 58 GLULEUPROVALILEAPRALILEVALTHRHISALAPROGLUVALPROPROVALAAR 77
DB 504 GAAGCGCCGATCATGACAAAG---GCGACGATGACGCCACCGTGCATCCACGTCGGT 448
QY 78 ARGAPRHISPROVALALYVALVAL-----LYSMETGLU--- 89
DB 447 CCGACTATCGCCCGGTGTGACGCTGAACGCGTGTCTGTGCATGGCGCATGAACGCG 388
QY 90 -----THRYALGLYUVALMETARG---LEUVALAARGLYVAL 101
DB 387 GACGTGAAGAGATTCACTGTGGCGGACAACTGTGTGCGGAATTGCCAAGGCATG 328
QY 102 GLUTRYGLINPHEPTRHPHEGLYGLINVALPROGLYGLMETLEARGVALARGGLU 121
DB 327 AAGGCTCATCTCTGGGTTCACAGCCCAATCCCGGTCACACCATTGAGGCACTAGAA 268
QY 122 GLYAPERTHRLIGLVALGLINPHESEARHISPROAERSELYMECPROHISAEN--- 140
DB 267 GGGGACAGCTCGTATCTTCTGTACCAAC-----AAATTGCGGACGCTTACC 220
QY 141 ---VALASPHENISIALAATHRGILYPROGLY-----GLYGLYALAGLU 154
DB 219 ACCGTTCATTTGGATGGCGGTGTTGCCGAGCGATGACGCGGTCCGCGGCTCAAC 160
QY 155 ALASERPHETRALAPROGLYHISTHRSERPHASEPHALYALALEUGINPROGLY 174
DB 159 CAGCTCATCATCAACTGCGGACGACCTTCTGTACGAGTTCCAGATGAAGCATACGCGG 100

```

[illegible]

Df		165	GAGTTTCCTGGTCCACCATTAAGAAGCCTTGCCGAGACACATCGTCGCAATCTCAC	224
Oy		131	AsnhiBProApsSerLyMetProHiaSn-----ValaSpneHiala-----	145
Df		225	AAC-----AAACTCACCCAGGAAGCCTTGCTCATCGATTGCATGATCCGT	272
Oy		146	-----AlathGlyProGlyGlyAlaSuAlaserPheThrala	159
Df		273	CAGTTCCGAAGTCCATGGCAGATGGAGACGACGAGATGTTACTCATGGCCAAATTAAC--	329
Oy		160	ProGlyHisThrSerThrPheSerPheLysAlaLeuGlnProGlyLeuTyrrValTyrlHis	179
Df		330	CCTGGAGAGACTTTTACTCTACAATTTACTGTTGAAAAGCCGGGGAACACATTCTCATCAT	389
Oy		180	CysAlaValAlaProValGlyMetHisIleAlaasnGlyMetTyrrGlyLeuIleVal	199
Df		390	-----GGACACTATGGCATGCACAGATACACTGCGCTATACGATCGTTATATGTG	440
Oy		200	GlutProLysGluGlyLeuPro-----LysValaAspLysGlyTyrrTyrrValMecGln	216
Df		441	GACCTGGCTAAAGAAAGAACGACGAGATTGAGATACATGGTGTGATTATCTCTTACTC	500
Oy		217	GlyAspPheTyrrThrLysGlyLysTyrrGlyGluGlnGlyLeuGlnProPheAspMet	235
Df		501	AGTACACTGCTGGCATGAGCTATTCCTCCCAAGACATCGTCTTTCT-----	548
Oy		236	GluLysAlaIleArg-----GluAspAlaGluTyrrValValPheasnGly	250
Df		549	TCCAAACCTTATGCCGTGATCGGTGAGCTCANAGCATATTGATAAATGGG	599
RESULT_34				
LOCUS		BG448142	NF106G08BC1F106E Elicited cell culture Medicago truncatula cDNA	
DEFINITION		BG448142	Clone NF106G08EC 5', mRNA sequence.	
ACCESSION		BG448142		
VERSION		BG448142.1	GI:13366923	
KEYWORDS			EST.	
SOURCE			Medicago truncatula (barrel medic)	
ORGANISM			Medicago truncatula	
			Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;	
			Spermatophyta; Magnoliophyta; eudicotyledons: core eudicots: rosids	
			; eurosoids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;	
			Medicago.	
REFERENCE			1 (bases 1 to 659) *	
AUTHORS			Torres-Jerez,I., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell,C.J.,	
			Flores,H.R., Imman,J.T., Weller,J.W. and May,G.D.	
TITLE			Expressed Sequence Tags from the Samuel Roberts Noble Foundation -	
JOURNAL			Center for Medicago Genomics Research	
COMMENT			Unpublished	
			Contact: Dixon RA	
			Plant Biology Division	
			The Samuel Roberts Noble Foundation	
			2510 Sam Noble Parkway, Ardmore, OK 73402, USA	
			Tel: 580 221 7302	
			Fax: 580 221 7380	
			Email: radixon@noble.org	
			Insert Length: 659 Std Error: 0.00	
			Plate: 106 row: G column: 08	
			Seq primer: TCACACAGCAAACAGCATATAC.	
FEATURES			Location/Qualifiers	
source			1..659	
			/organism="Medicago truncatula"	
			/mol_type="mRNA"	
			/db_xref="taxon:3880"	
			/cclone="NF106G08EC"	
			/tissue_type="Cell cultures derived from root tissues"	
			/dev_stage="Cell suspensions were subcultured every 14	
			days. Cells were induced six days after subculture"	
			/clone_lib="Elicited cell culture"	
			/note="Vector: Lambda Zap; Cells were induced with yeast	
			cell wall extracts equivalent to 50ug/ml glucose in the	
			final concentration. Samples were taken at 0.5, 1, 12 and	

24 hours after induction. Equal amounts of RNA from each time point were pooled and used for mRNA isolation."

## ALIGNMENT SCORES:

Prod. No.:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
Score: 0.000314	659	59	29	88	50	10
Percent Similarity: 139.50						
Best Local Similarity: 38.94%						
Query Match: 26.11%						

US-10-088-045-2 (1-502) x BG448142 (1-659)

```

QY 24 SerAnGlnAlaAspLysAlaGlnProLysSerSerThrValAspAlaAlaLys 43
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 86 TCTAGACATGCTTCACTTCACTAAGCCCTAAGGCGACCAAAATTATTGTTCTATG 145
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 44 ThrAlaAsnAlaAspAlaAlaSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 63
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 146 TTTCTTTCT-----CATTTGCTGAATTTCCAAAGCTGA 181
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 63 palAlaIleValThrHisAlaProGluValProProValAspArgAspHisProAlaLys 83
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 182 GCGTGGCTTACAT-----197
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 83 sValValValLysMetGlnThrValGlnLysValMetArgLysAlaAspGlyValGlnLys 103
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 198 -----TATTAATGGAG--GTGAAGTATGATTATAGATCCCTGATGTTATAGAA 247
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 103 rGlnPheTrpThrPheGlyGlyGlnValProGlyGlnMetIleArgValArgGlnGlnLys 123
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 248 ACTGTTTATACCATTAATGAAAGACTCCAGAGCTTCACTTCAAGCAGAGAGGCTGA 307
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 123 pThrIleGlnValGlnPheSerAsnHisProAspSerLysMetProHisAsnValAspPh 143
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 308 TACTGTTGATGTTCAACTTAACAC-----AAATGCTCAGAAAACCTGCCAT 358
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 143 eHis-----AlaIleThrGlyProGlyGlyVal 153
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 359 CCATTTGGCATGATATTAGACAGATTGGAACTCTTGGTTGATGGAACAGAGAGTGTG 418
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 153 aGlnAlaSerPheThrAlaProGlyHisThrSerThrPheSerPheLysAlaGlnInr 173
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 419 TCATATGCTTATA---CTACCTGGAGACACCTTTGTTATCGATTGTTGATAGGCC 475
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 173 oGlyLeuTyValTyHisCysAlaValAlaProValGlyMetHisIleAlaAsnGlyMe 193
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 476 TGTGATCATATCTTATCAT-----GCTCAGTATGGAATGCAAGAGAGAGT 526
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 193 tTyGlyLeuIleLeuValGlnProLysGlnGlyLeuPro--LysValAspLysGlnLys 212
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 527 ATATGAGATGATGCTGTGCACTTAATGACCTGAACTTTTCTTATGCTTATAG 586
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 212 rTyValMetGlnGlnLysPheTrpThrLysGlyLysTyGlyGlnGlnGlnGlnGln 231
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 587 AAGCATTAATTTGATGATGTTGATGATGATGATGATGATGATGATGATGATGAT 643
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 232 -----ProPhe 233
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 644 GTCTGCAATCCCTTTC 659
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

## RESULT 35

BF644576 690 bp mRNA linear EST 20-DEC-2000  
 LOCUS NP015099.1079 Elicited cell culture Medicago truncatula cDNA  
 DEFINITION clone NF015H09EC 5', mRNA sequence.  
 ACCESSION BF644576  
 VERSION BF644576.1 GI:11909705  
 KEYWORDS EST.  
 SOURCE Medicago truncatula (barrel medic)  
 ORGANISM Medicago truncatula

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosoid I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae; Medicago.

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

1 (bases 1 to 690)  
 Torres-Jerez, I., Scott, A.D., Harris, A.R., Gonzales, R.A., Bell, C.J., Flores, H.R., Imman, J.T., Weller, J.W. and May, G.D.  
 Expressed Sequence Tags from the Samuel Roberts Noble Foundation - Center for Medicago Genomics Research  
 Unpublished  
 Contact: Dixon RA  
 Plant Biology Division  
 The Samuel Roberts Noble Foundation  
 2510 Sam Noble Parkway, Ardmore, OK 73402, USA  
 Tel: 580 221 7302  
 Fax: 580 221 7380  
 Email: radixon@noble.org  
 Insert Length: 690 Std Error: 0.00  
 Plate: 015 row: H column: 09  
 Seq primer: TCACAGAGAAACAGCTATGAC.  
 Location/Qualifiers

## FEATURES

## source

1..690  
 /organism="Medicago truncatula"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:3880"  
 /clone="NF015H09EC"  
 /tissue="Cell cultures derived from root tissues"  
 /dev\_stage="Cell suspensions were subcultured every 14 days. Cells were induced six days after subculture"  
 /note="Vector: Lambda Zap; Cells were induced with yeast cell wall extracts equivalent to 50ug/ml glucose in the final concentration. Samples were taken at 0.5, 1, 12 and 24 hours after induction. Equal amounts of RNA from each time point were pooled and used for mRNA isolation."

## BASE COUNT

## ORIGIN

202 a 136 c 135 g 217 t

## ALIGNMENT SCORES:

Prod. No.:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
Score: 0.000335	690	59	29	88	50	10
Percent Similarity: 139.50						
Best Local Similarity: 38.94%						
Query Match: 26.11%						

US-10-088-045-2 (1-502) x BF644576 (1-690)

```

QY 24 SerAnGlnAlaAspLysAlaGlnProLysSerSerThrValAspAlaAlaLys 43
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 87 TCTAGACATGCTTCACTTCACTAAGCCCTAAGGCGACCAAAATTATTGTTCTATG 146
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 44 ThrAlaAsnAlaAspAlaAlaSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 63
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 147 TTTCTTTCT-----CATTTGCTGAATTTCCAAAGCTGA 182
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 63 palAlaIleValThrHisAlaProGluValProProValAspArgAspHisProAlaLys 83
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 183 GCGTGGCTTACAT-----198
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 83 sValValValLysMetGlnThrValGlnLysValMetArgLysAlaAspGlyValGlnLys 103
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 199 -----TATTAATGGAG--GTGAAGTATGATTATAGATCCCTGATGTTATAGAA 248
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 103 rGlnPheTrpThrPheGlyGlyGlnValProGlyGlnMetIleArgValArgGlnGlnLys 123
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 249 ACTGTTTATACCATTAATGAAAGACTCCAGAGCTTCACTTCAAGCAGAGGCTGA 308
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 123 pThrIleGlnValGlnPheSerAsnHisProAspSerLysMetProHisAsnValAspPh 143
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 309 TACTGTTGATGTTCAACTTAACAC-----AAATGCTCAGAAAACCTGCCAT 359
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 143 eHis-----AlaIleThrGlyProGlyGlyVal 153
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Db 360 CCATTGGCATGTATTAGACAGATTGAACTCTGTTTGGATGGAACAGAAAGAGTGTTC 419  
 153 acguasaserpethrAlaProGlyHisThrSerThrPheSerPheLysAlaLeuGlnP 173  
 Db 420 TCACATCTCTATA---CTACCTGGAGACACCTTTGTTTATCGATTGTTGTTGATGGCC 476  
 173 oclyleuTyrrValTyrrHisCysAlaValAlaProValGlyMetHisIleAlaAsnGly 193  
 477 TGGTACATATCTTTATCAT-----GCTCACTATGAGATGCAAGGAGACAGAGT 527  
 193 tTyrrGlyLeuIleLeuValGluProLysGluGlyLeuPro--LysValAspLysGlu 212  
 528 ATATGGAATGATTGTTGTCGACACCTATATGACCTGAACTTTTCTTATGACTTGTATG 587  
 212 tTyrrValMetGlnGlyAspPheTyrrThrLysGlyLysTyrrGlyGluGlnGlyLeu 231  
 588 AAGCATATTATTTGATGATGTTGGTACCATAGAGTACTTAT--GAACAAATGCTGATGATT 644  
 Qy 232 -----ProPhe 233  
 Db 645 GTCTGCATCCCTTTT 660

RESULT 36  
 AV826844 571 bp mRNA linear EST 01-APR-2002  
 LOCUS AV826844 RAFL9 Arabidopsis thaliana cDNA clone RAFL09-07-M21 5',  
 DEFINITION mRNA sequence.  
 AV826844  
 VERSION AV826844.1 GI:19868904  
 SOURCE EST.  
 ORGANISM Arabidopsis thaliana (thale cress)  
 Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids  
 ; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 1 (bases 1 to 571)  
 REFERENCE  
 AUTHORS Seki, M., Narusaka, M., Ishida, J., Kamiya, A., Satou, M., Nakajima, M.,  
 Oono, Y., Sakurai, T., Carninci, P., Kawai, J., Itoh, M., Ishii, Y.,  
 Arakawa, T., Shibata, K., Shinagawa, A., Muramatsu, M., Hayashizaki, Y.  
 and Shinozaki, K.  
 Large scale analysis of Arabidopsis full-length cDNA (2002b)  
 TITLE  
 JOURNAL Unpublished  
 COMMENT Contact: Motoaki Seki  
 Plant Functional Genomics Research Group  
 RIKEN Genomic Sciences Center  
 3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan  
 Tel: 81-298-36-4359  
 Fax: 81-298-36-9060  
 Email: msek@rcc.riken.go.jp  
 An Arabidopsis full-length cDNA library was constructed essentially  
 as reported previously (Seki et al., 1998). cDNA cleaved with BamHI  
 and XhoI was ligated to modified Lambda FLC-1 vector (Carninci et  
 al., submitted for publication) digested with BamHI and SalI. This  
 clone is in a modified Bluescript vector. Please visit our web  
 site ([http://www.gsc.riken.go.jp/e/planet/index\\_e.html](http://www.gsc.riken.go.jp/e/planet/index_e.html)) for further  
 details.

FEATURES  
 source location/Qualifiers  
 1..571  
 /organism="Arabidopsis thaliana"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:3702"  
 /clone="RAFL09-07-M21"  
 /dev\_stage="plants at various developmental stages from  
 germination to mature seeds"  
 /lab\_host="DH10B"  
 /clone\_1ib="RAFL9"  
 /note="Site 1: BamHI; Site 2: SalI; subjected to  
 dehydration (1, 2, 5, 10, 24 hr) and cold (1, 2, 5, 10, 24  
 hr) treatments"  
 145 a 133 c 154 g 136 t 3 others

BASE COUNT 145 a 133 c 154 g 136 t 3 others  
 ORIGIN

Alignment Scores:  
 Pred. No.: 0.000291 Length: 571  
 Score: 139.00 Matches: 39  
 Percent Similarity: 45.14% Conservative: 26  
 Best Local Similarity: 27.08% Mismatches: 47  
 Query Match: 5.29% Indels: 32  
 DB: 9 Gaps: 7

US-10-088-045-2 (1-502) x AV826844 (1-571)  
 Qy 101 ValGluTyrrGlnPheTrp-----ThrPheGlyGly 110  
 105 GTGAGGTACAGATATTGTCGCCGAGCTGCAAGAGCGCCGTTATATGACCTCAACGGC 164  
 111 GlnValProGlyGlnMetIleArgValArgGluGlyAspThrIleGluValGlnPheSer 130  
 165 GAGTTTCTCTGTCGCCACCATTAAGCTTCCGCGAGACACCATGCTGCTCATCTCAC 224  
 Qy 131 AsnHisProAspSerLysMetProHisAsn-----ValAspPheHisAla----- 145  
 225 AAC-----AACTCACACCGAAGCGCTTGTCATCTCATTTGGCATGGAATCCGT 272  
 Qy 146 -----AlaThrGlyProGlyGlyGlyAlaGluAlaSerPheThrAla 159  
 273 CAGTTCGGAAGTCATGCGCAGATGAGCAGCAGAGTTCATCTCAATGCGCAATTAAAC-- 329  
 160 ProGlyHisThrSerThrPheSerPheLysAlaLeuGlnProGlyLeuTyrrValTyrrHis 179  
 330 CCGAGAGACACTTTTACCATATTTCACCTGTTGAAAAGCCGGGAACACATTTCTACCAT 389  
 180 CysAlaValAlaProValGlyMetHisIleAlaAsnGlyMetTyrrGlyLeuIleLeuVal 199  
 390 -----GCACTATAGGATGAGATGAGATGAGTGGGCTTATAGGATCGTTGATGG 440  
 Qy 200 GluProLysGluGlyLeuPro-----LysValAspLysGluTyrrTyrrValMetGln 216  
 441 GACGTGGCTTAAGAAAGAGCGAGATTCAGATTCAGATGCTGATTTATCTTATCTTAC 500  
 Qy 217 GlyAspPheTyrr 220  
 Db 501 ACTGACTGCTGG 512

RESULT 37  
 CA553614 489 bp mRNA linear EST 19-NOV-2002  
 LOCUS CA553614/c 489 bp mRNA linear EST 19-NOV-2002  
 DEFINITION C0874E11-5N NIA Mouse Blastocyst cDNA library (long) Mus musculus  
 cDNA clone NIA:C0874E11 IMAGE:30032122 5', mRNA sequence.  
 CA553614  
 VERSION CA553614.1 GI:25097868  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 489)  
 REFERENCE  
 AUTHORS Piao, Y., Kargul, G.J., Dudekula, D.B., Qian, Y., Tanaka, T., Luo, A. and  
 Ko, M.S.H.  
 TITLE  
 JOURNAL Systematic Analyses of NIA Mouse Blastocyst cDNA Library (Long)  
 COMMENT Unpublished  
 CONTACT Dawood B. Dudekula  
 Laboratory of Genetics  
 National Institute on Aging/National Institutes of Health  
 333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA  
 Email: [cdna@igsun.grc.nia.nih.gov](mailto:cdna@igsun.grc.nia.nih.gov)  
 Plate: C0874 row: B column: 11  
 Seq primer: M13 Reverse  
 High quality sequence stop: 489  
 POLYA-No.

FEATURES  
 source location/Qualifiers  
 1..489  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strains="C57BL/6J"

```

BASE COUNT
ORIGIN

99 a      144 c      144 g      102 t

/db_xref="niaEST: C0874E11-SN"
/db_xref="taxon:10090"
/clone="NIA: C0874E11 IMAGE:30032122"
/tissue_type="Blastocyst"
/dev_stage="3.5-dpc"
/lab_host="DH10B"
/clone_1ib="NIA Mouse Blastocyst cDNA library (long)"
/note="Vector: pSPORT1 (Invitrogen); Site_1: SalI; Site_2:
NotI; Mouse cDNA project by the Laboratory of Genetics,
National Institute on Aging (NIA), Intramural Research
Program, NIH (http://lgsun.grc.nia.nih.gov/cDNA). This is
a long-transcript enriched cDNA library (Ref. Genome Res.
11: 1553-1558 (2001). [PMID: 11544199]). Total RNAs were
extracted from a pool of 20 Blastocysts. Double-stranded
cDNAs were synthesized with an Oligo(dT) primer.
Invitrogen:
5'-GACAGTACTTCATGATCGCAGCGCGCCGCTTTTTTTTTTTT-3'] from
0.2 ug of total RNA, treated with T4 DNA polymerase, and
purified by ethanol-precipitation. The cDNAs were ligated
to lone-linker LL-Sal4, purified by phenol/chloroform, and
separated from free linkers by Centricon 100. Then, the
cDNAs were amplified by long-range high fidelity PCR using
Ex Tag polymerase (Takara) with a primer Sal4-S. The
products were purified by phenol/chloroform and Centricon
100. The cDNAs were digested with SalI and NotI enzymes
and cloned into SalI/NotI site of pSPORT1 plasmid vector.
The DH10B E. coli host was transformed with the ligation
mixture by the standard kb method. The average
insert size is about 2.2 kb. The library was constructed
by Yulan Piao (NIA)."

```

Alignment Scores:			
Pred. No.:	0.000266	Length:	489
Score:	138.50	Matches:	35
Percent Similarity:	47.50%	Conservative:	22
Best Local Similarity:	29.17%	Mismatch:	50
Query Match:	5.27%	Indels:	13
DB:	14	Gaps:	5

  

US-10-088-045-2 (1-502) x CA553614 (1-489)	
Qy	92 GluIysValMetArg---LeuAlaSpIValGluIYrgInPheTrpThrPheGlyGly 110
Db	360 GAGCGCGGTCTGCTCCGTAATTGCGGAGGATGAAGCTCACTTATGAGGGGTATTAAGCT 301
Qy	111 GluValProGlyGlnMetIleArgValArgGluGlyAspThrIleGluValGlnPheSer 130
Db	300 CAATGCGCGGGCCGACCATTTAGGCGCTCGAAGCGCAGAAGCTCGCATTTGGTCACG 241
Qy	131 AsnHisProAspSerIlyMetProHis-----AsnValAspPheHisAlaAlaThrGly 148
Db	240 AAC-----AAACTGCCGAGACAGACGACCATTCATTGGCAGCGTGTGCTTGTG 193
Qy	149 ProGly-----GlyGlyAlaGluAlaSerPheThrAlaProGlyHisThr 163
Db	132 CCCAAGCGCATGAGACGGCGTGTGTGACTCAATCAGCCCCACATTTCCCGGCGAAGACC 133
Qy	164 SerThrPheSerPheIlyAlaLeuGlnProGlyLeuIYrgValIYrHisCysAlaValAla 183
Db	132 TTCGTGTACGACTTCAGATGAAGACACACGGGAGCTTCATGTATCAT---CCGATTCG 76
Qy	184 ProValGlyMetHisIleLeuAsnGlyMetIYrgIlyLeuIleLeuValGlnProIyGln 203
Db	75 GACGAATATGCTCCAGATGCGGATGCGATGATGAGATGATGATGCTGATCCGCGCGAT 16

  

RESULT 38			
LOCUS	BF650165	645 bp	mRNA linear
DEFINITION	NF088G08EC1F1066 Elicited cell culture Medicago truncatula cDNA		
ACCESSION	clone NF088G08EC 5', mRNA sequence.		
	BF650165		

VERSION	BF650165.1	GI:11915295
KEYWORDS	EST.	
SOURCE	Medicago truncatula (barrel medic)	
ORGANISM	Medicago truncatula	
	Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eucosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae; Medicago.	
REFERENCE	1 (bases 1 to 645)	
AUTHORS	Torres-Jerez, I., Scott, A. D., Harris, A. R., Gonzales, R. A., Bell, C. J., Flores, H. R., Imman, J. T., Weller, J. W. and May, G. D.	
TITLE	Expressed Sequence Tags from the Samuel Roberts Noble Foundation - Center for Medicago Genomics Research	
JOURNAL	Unpublished	
COMMENT	Contact: Dixon RA Plant Biology Division The Samuel Roberts Noble Foundation 2510 Sam Noble Parkway, Ardmore, OK 73402, USA Tel.: 580 221 7302 Fax: 580 221 7380 Email: radixon@noble.org Insert Length: 645 Std Error: 0.00 Plate: 088 row: G column: 08 Seq primer: TCACACGGAACGACGTATGAC.	
FEATURES	Location/Qualifiers	
SOURCE	1..645	

```

/tissue_type="cell cultures derived from root tissues"
/dev_stage="Cell suspensions were subcultured every 14
days. Cells were induced six days after subculture"
/clone_id="Elcited cell culture"
/note="Vector: Lambda Zap; Cells were induced with yeast
cell wall extracts equivalent to 50ug/ml glucose in the
final concentration. Samples were taken at 0.5, 1, 12 and
24 hours after induction. Equal amounts of RNA from each
time point were pooled and used for mRNA isolation."
BASE COUNT      189 a      116 c      136 g      203 t      1 others
ORIGIN
Alignment Scores:
Pred. No.:          0.000438      Length:        645
Score:             139.00         Matches:        45
Percent Similarity: 40.96%       Conservative:   23
Best Local Similarity: 27.11%     Mismatches:    58
Query Match:       5.26%         Indels:        40
DB:                10           Gaps:          8
US-10-088-045-2 (1-502) x BF650165 (1-645)

QY      100 G|yValG|utyrgInPheTr----- 106
            |||||
            ::::
Db       90 GGTTGTAGACATTATAAATGGAGGTGAAGTATTAAGATCCCTGATTGTTATAG 149
                        |||||
QY      107 -----ThrPheGLyGLINValProGLyImeCILEArGvalArGluGLy 122
                        |||||
                        ::::
Db       150 AAActGTGTAAACATTATTAAGAAAGACTCAAGACCTACCATTCAAGACAAGAGGCT 209
                        |||||
QY      123 AspThrIleGLUVALglnPheSerAsnHisProAspserLysMeCProHISAsNValAsp 142
            |||||
            ::::
            ::::
Db       210 GATAcTGtTGTAgTTGAAGTTAACAC-----AAATTCGCACAGAAGAACTTCCC 260
            |||||
QY      143 PheHis-----AlaIalThrGLyProGLyGLy 152
            |||||
Db       261 ATCATTTGGCAGTGTATTAGACAGATTGGAACTCCTTGTTAGTAGAAGAGAGAGCTG 320
            |||||
QY      153 AlAGlualaserPheThrAlarProGLyHisThrSerThrPheSerPheLysAlaleuGLN 172
            ::::
            |||||
            ::::
Db       321 TCTCAATGTCTTA--CTACCTCGAGACACCTTGTGTTATGCATTTGTTGTGATAGG 377
            |||||
QY      173 ProGLyLeuTyGValTYrHiSCyalaValalarProValGLyMeHIsILaIalsmgly 192

```

Db 378 CTTGATACATCTTATATAT-----GCTCACTATGAAATGCAAGGAGGACGGA 428  
 Qy 193 MetTyrGlyLeuIleValGluProLysGluGlyLeuPro---LysValAspLysGlu 211  
 Db 429 GATATGGAATGATTCGTGGACCTATGACCCCTGAACCTTTTCTTATGACCTTGTAT 488  
 Qy 212 TyrTyrValMetGlnGlyAspPheTyrThrLysGlyLysTyrGlyGlnGlnGlyLeuIn 231  
 Db 489 AGAAGCATATTTTGAATGATTCATGACATAGAGTACTTAT---GAACAACTGCTAGA 545  
 Qy 232 -----ProPhe 233  
 Db 546 TTGCTGCAATCCCTTTT 563

RESULT 39  
 BU635352 737 bp mRNA linear EST 23-SEP-2002  
 LOCUS BU635352  
 DEFINITION 020H08 Infected Arabidopsis leaf Arabidopsis thaliana cDNA, mRNA  
 sequence.  
 ACCESSION BU635352.1 GI:23302607  
 VERSION BU635352.1  
 KEYWORDS EST.  
 SOURCE Arabidopsis thaliana (thale cress)  
 ORGANISM Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids  
 / eustosids II; Brassicales; Brassicaceae; Arabidopsis.  
 REFERENCE 1 (bases 1 to 737)  
 AUTHORS Lundgaard, M., Emmersen, J., Nielsen, K.L., Wilson, I., Somerville, S.  
 and Weindler, K.G.  
 TITLE EST sequencing of Erysiphe cichoracearum infected Arabidopsis  
 plants  
 JOURNAL Unpublished  
 COMMENT Contact: Karen G. Weindler  
 Institut for bioteknologi  
 Aalborg Universitet  
 Sohngaardsholmsvej 49, 9000 Aalborg, Denmark  
 Tel: +45 96358467  
 Fax: +45 98141808  
 Email: kgy@bio.au.dk.  
 FEATURES  
 source  
 1. 737  
 Location/Qualifiers  
 /organism="Arabidopsis thaliana"  
 /mol\_type="mRNA"  
 /strain="Columbia"  
 /db\_xref="taxon:3702"  
 /dev\_stage="Plant 3 weeks old, three days post infection"  
 /clone\_lib="Infected Arabidopsis leaf"  
 /note="Organ: Leaf; Vector: pBluescript; Mixed cDNA  
 library of Arabidopsis and E. cichoracearum infected leaf  
 from three weeks old Arabidopsis plants. Plants were  
 harvested 3 days after infection and mRNA oligo dt  
 selected."  
 BASE COUNT 208 a. 164 c 164 g 201 t  
 ORIGIN  
 Alignment Scores: 0.00067 Length: 737  
 Pred. No.: 137.00 Matches: 67  
 Score: 37.55% Conservative: 28  
 Percent Similarity: 37.55% Mismatches: 96  
 Best Local Similarity: 26.48% Indels: 62  
 Query Match: 5.22% Gaps: 13  
 DB: 13

US-10-088-045-2 (1-502) x BU635352 (1-737)  
 Qy 107 ThrPheGlyGlyGlnValProGlyGlnMetIleArgValArgGluGlyAspThrIleGlu 126  
 Db 25 ACAATCAATGTAAGTTCAGGTCCCAACCATTAAGCTCAACAAAGGTGACACCATCGTT 84  
 Qy 127 ValGlnPheSerAsnHisProAspSerLysMetProHisAsnValAspPheHis----- 144

Db 85 GTGAGCTCAAGAT-----AGCTTCATGACTGAAAATGCTGCTCCATGGCAT 135  
 Qy 145 -----AlaIleThrGlyProGlyGlyGlyAlaGluAlaSerPheThrAla--- 159  
 Db 136 GGAATCGACAGATTGGGACTCCATGTTGATGAGTGAAGGTGTACTCAATGTCCA 195  
 Qy 160 -----ProGlyHisThrSerThrPheSerPheLysAlaLeuGlnProGlyLysLeuTyrVal 177  
 Db 196 ATCTTCCTCGAAGAGCTTCATTACCAATTGTGCTGATGAGGCTGTCATACATG 255  
 Qy 178 TyrHisGlyAlaValAlaProValGlyMetHisIleAlaAsnGlyMetTyrGlyLeuIle 197  
 Db 256 TATCAT-----TCACACTATGGGATGACAGAGAGTGTGATTAATCGGAATGATT 306  
 Qy 198 LeuValGluProLysGluGlyLeuProLys-----ValAspLysGluTyrTyrVal 214  
 Db 307 CAGATTTCCTCCCGGCACAGAGCCGAACCGTTATACATGATATGACCGAACTTT 366  
 Qy 215 MetGlnGlyAspPheTyrThrLysGlyLysTyrGlyGlu---GlnGlyLeuGln----- 231  
 Db 367 TTGTTAAGTATGTTGTTATCACAAAAGTATGTCAGAGAAACCAACCGGTTAGCATCATA 426  
 Qy 232 ProPheAspMetGluLysAlaIleArgGluAspAlaGluTyrValVal----- 247  
 Db 427 CCCTTC-----AAGTGGTCCGTGAGCCCAATCGCTTATGATACAAAGAGAGA 477  
 Qy 248 ---PheAsnGlySer-----ValGlyAlaLeuThr 256  
 Db 478 AGATTCAAGCTCTCAAAACCAACCTACCACTCTCCAGCTTAGTGTGAGATGTATAC 537  
 Qy 257 GlyLysLeuAla-----LeuLysAlaLysValGlyGluThrValArg 270  
 Db 538 GTCTCAACCGTGAATGTTGCGGTTTATTTTTCAGGTGATCCCGAAGAACATACAG 597  
 Qy 271 LeuPheValGlyAsnGlyGlyProAsnLeuThrSerSerPheHisValIleGlyLysIle 290  
 Db 598 CTTCGAATAGTACTGATGAGCGCTCTCTGCTCGTCAAGTTCCCAATC----- 645  
 Qy 291 PheAspLysValHisPheGluGlyGlyLysGluLysHisAsnIleGlnThrLeu 310  
 Db 646 -----GAGGACACVAAATTTG-----ACAGTT 666

RESULT 40  
 AV826343 612 bp mRNA linear EST 01-APR-2002  
 LOCUS AV826343  
 DEFINITION AV826343 RAF18 Arabidopsis thaliana cDNA clone RAF108-13-A06 5',  
 mRNA sequence.  
 ACCESSION AV826343.1 GI:19868403  
 VERSION AV826343  
 KEYWORDS EST.  
 SOURCE Arabidopsis thaliana (thale cress)  
 ORGANISM Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids  
 / eustosids II; Brassicales; Brassicaceae; Arabidopsis.  
 REFERENCE 1 (bases 1 to 612)  
 AUTHORS Seki, M., Narusaka, M., Ishida, Y., Kamiya, A., Satou, M., Nakajima, M.,  
 Oono, Y., Sakurai, T., Carninci, P., Kawai, J., Itoh, M., Ishii, Y.,  
 Arakawa, T., Shibata, K., Shinagawa, A., Muramatsu, M., Hayashizaki, Y.  
 and Shinozaki, K.  
 TITLE Large scale analysis of Arabidopsis full-length cDNA (2002b)  
 JOURNAL Unpublished  
 COMMENT Contact: Motoaki Seki  
 Plant Functional Genomics Research Group  
 RIKEN Genomic Sciences Center  
 3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan  
 Tel: 81-298-36-4359  
 Fax: 81-298-36-9060  
 Email: mseki@rtc.riken.go.jp

An Arabidopsis full-length cDNA library was constructed essentially as reported previously (Seki et al., 1998). cDNA cleaved with BamHI and XhoI was ligated to modified lambda FL-C-1 vector (Carninci et al., submitted for publication) digested with BamHI and SalI. This clone is in a modified Bluescript vector. Please visit our web site ([http://www.gsc.riken.go.jp/e/plant/index\\_e.html](http://www.gsc.riken.go.jp/e/plant/index_e.html)) for further details.

## FEATURES

## source

1.612 Location/Qualifiers

/organism="Arabidopsis thaliana"

/mol\_type="mRNA"

/db\_xref="taxon:3702"

/clone="RAF08-13-A06"

/dev\_stage="rosette plants"

/lab\_host="DH10B"

/clone\_id="RAF08"

/note="Site 1: BamHI; Site 2: SalI; subjected to dehydration-treated (1, 2, 5, 10, 24 hr)"

BASE COUNT 160 a 140 c 165 g 144 t 3 others

## ORIGIN

## Alignment Scores:

Prod. No.:	0.000586	Length:	612
Score:	136.50	Matches:	46
Percent Similarity:	44.07%	Conservative:	32
Best Local Similarity:	25.99%	Mismatches:	60
Query Match:	5.20%	Indels:	39
DB:	9	Gaps:	10

US-10-088-045-2 (1-502) x AV826343 (1-612)

```
OY 101 ValGluTyrGlnPheTrp-----ThrPheGlyGly 110
DB 105 GTGAGAGTACAGTATGTCGCCGACTGCMAAGAGCGCCGTTATGACCGTCAACGGC 164
OY 111 GluValProGlyGlnMetIleArgValArgGluGlyAspThrIleGluValGlnPheSer 130
DB 165 GAGTTTCCTGCTCCACCATTAAGCCTTGCCGCGAGACACCATCGTCGTCATCTCACC 224
OY 131 AsnHisProAspSerIleMetProHisAsn-----ValAspPheHisAla----- 145
DB 225 AAC-----AAATCACCACCGAAGCGCTTGTCTCATCTGCAATGGATGATCGGT 272
OY 146 -----AlaThrGlyProGlyGlyValAlaGluAlaSerPheThrAla 159
DB 273 CAGTTCCGAACTCCATGGCGAGATGAGACAGAGAGTACTCAATGCCAATTAC--- 329
OY 160 ProGlyHisThrSerThrPheSerPheIleValLeuGlnProGlyLeuTyrValTyrHis 179
DB 330 CCTGGAGAGACTTTTACTTACATTTCTCACTGTGAAAAGCCGCGAGACACATTTTACCAT 389
OY 180 CysAlaValAlaProValGlyMetHisIleAlaAsnGlyMetTyrGlyLeuIleLeuVal 199
DB 390 -----GGACACTATGGCATGCAGATCAGATCGGCGTATAGCGATCGTGTATTGTG 440
OY 200 GluProIleGluGlyLeuPro-----LysValAspLysGluTyrTyrValMetGln 216
DB 441 GACGTGGCTAAAGGAAAGAGGAGATGAGATACATGATGATGATTAATCTTACTC 500
OY 217 GlyAspPheTyrThrLysGly--LysTyrGlyGluGlnGlyLeuGlnProPheAspMet 235
DB 501 AGTACTGTGGTGCATGAGCTATTCCTCCCAAGAACTCGGCTTCTC----- 548
OY 236 GluLysAlaIleArg-----GluAspAlaGluTyrValValPheAsnGly 250
DB 549 TCCAAACTATGCGCTGATCGTGAAGCTCAGANCATATTGATTAATGGG 599
```

Search completed: September 5, 2003, 13:17:57  
Job time : 3101 secs